

Urban land-use classification using machine learning classifiers: comparative evaluation and post-classification multi-feature fusion approach

Yashon O. Ouma, Amantle Keitsile, Boipuso Nkwaе, Phillimon Odirile, Ditiro Moalafhi & Jiaguo Qi

To cite this article: Yashon O. Ouma, Amantle Keitsile, Boipuso Nkwaе, Phillimon Odirile, Ditiro Moalafhi & Jiaguo Qi (2023) Urban land-use classification using machine learning classifiers: comparative evaluation and post-classification multi-feature fusion approach, European Journal of Remote Sensing, 56:1, 2173659, DOI: [10.1080/22797254.2023.2173659](https://doi.org/10.1080/22797254.2023.2173659)

To link to this article: <https://doi.org/10.1080/22797254.2023.2173659>



© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 22 Feb 2023.



Submit your article to this journal



Article views: 5063



View related articles



View Crossmark data



Citing articles: 35 View citing articles

Urban land-use classification using machine learning classifiers: comparative evaluation and post-classification multi-feature fusion approach

Yashon O. Ouma ^a, Amantle Keitsile^a, Boipuso Nkwa^a, Phillipon Odirile^a, Ditiro Moalafhi^b and Jiaguo Qi ^c

^aDepartment of Civil Engineering, University of Botswana, Gaborone, Botswana; ^bFaculty of Natural Resources, BUAN, Gaborone, Botswana; ^cCenter for Global Change and Earth Observations, Michigan State University, East Lansing, Michigan, USA

ABSTRACT

Accurate spatial-temporal mapping of urban land-use and land-cover (LULC) provides critical information for planning and management of urban environments. While several studies have investigated the significance of machine learning classifiers for urban land-use mapping, the determination of the optimal classifiers for the extraction of specific urban LULC classes in time and space is still a challenge especially for multitemporal and multisensor data sets. This study presents the results of urban LULC classification using decision tree-based classifiers comprising of gradient tree boosting (GTB), random forest (RF), in comparison with support vector machine (SVM) and multilayer perceptron neural networks (MLP-ANN). Using Landsat data from 1984 to 2020 at 5-year intervals for the Greater Gaborone Planning Area (GGPA) in Botswana, RF was the best classifier with overall average accuracy of 92.8%, MLP-ANN (91.2%), SVM (90.9%) and GTB (87.8%). To improve on the urban LULC mapping, the study presents a post-classification multiclass fusion of the best classifier results based on the principle of feature in-feature out (FEI-FEO) under mutual exclusivity boundary conditions. Through classifier ensemble, the FEI-FEO approach improved the overall LULC classification accuracy by more than 2% demonstrating the advantage of post-classification fusion in urban land-use mapping.

ARTICLE HISTORY

Received 19 August 2022
Revised 21 January 2023
Accepted 24 January 2023

KEYWORDS

Urban land-use land-cover;
Gradient tree boosting;
Random forest; Support
vector machine; Multilayer
perceptron neural networks;
Post-classification feature
fusion

Introduction

Within the urban ecosystems, the determination and analysis of land-use and land-cover (LULC) and LULC change analysis plays an important role in providing the critical input in the decision-making process for environmental planning and ecological management (Dwivedi et al., 2005; Fan et al., 2007). For urban LULC mapping and change detection, remote sensing data provides the optimal spatial and temporal data sources. However, the extraction of urban features is often a challenging task due to the high degree of interactions and complexities within the features in terms of their spectral, spatial and textural properties (Blaschke et al., 2014). Due to these factors, the applications of traditional pixel-based classifiers in urban LULC mapping often lead to unsatisfactory results (Johnson & Xie, 2013; Myint et al., 2011).

To overcome the drawbacks in pixel-based classifications, Blaschke et al., 2014 proposed the geographic object-based image analysis (GEOBIA) focusing on the segmentation of very high-spatial resolution (VHR) image data. Using GEOBIA segmentation, pixels are grouped into similar and semantically independent image segments or objects for feature extraction and classification. For VHR image data, GEOBIA has been reported to perform better than pixel-based

approaches in urban LULC mapping (Blaschke et al., 2014; Drägut et al., 2010; Johnson & Xie, 2013; Jozdani et al., 2018). GEOBIA segmentation approach does not however yield good results for medium- and low-resolution remote sensing data.

At medium spatial resolutions, therefore, methods comprising of unsupervised algorithms, parametric supervised and machine learning methods have been widely used for LULC mapping (Friedl & Brodley, 1997; Halder et al., 2011; Li et al., 2016; Orieschnig et al., 2021; Waske & Braun, 2009; Wu et al., 2019). The supervised classifiers comprise of maximum likelihood classifier, Mahalanobis distance, k-nearest neighbors (kNN), support vector machine (SVM), random forest (RF), decision trees (DT), spectral angle mapper (SAM), fuzzy logic, fuzzy adaptive resonance theory-supervised predictive mapping (Fuzzy-ARTMAP), radial basis function (RBF), artificial neural networks (ANN) and naive Bayes (NB) (Ma et al., 2019; Shih et al., 2019). The unsupervised classifiers include, among others, fuzzy c-means, k-means algorithm, affinity propagation clustering algorithm and ISODATA techniques (Maxwell et al., 2018).

Urban LULC mapping is data intensive requiring both current and historical remote sensing data. To improve on the urban LULC mapping, machine learning (ML) and artificial intelligence classifiers have

been preferred (Mao et al., 2020; Lefulebe et al. (2022). In general, the application of ML algorithms for LULC mapping has attracted considerable research interests (Maxwell et al., 2018; Wang et al., 2022). This is mainly because ML algorithms do not require hypotheses on the input data distribution and tend to yield better results than the traditional parametric classifiers (Jozdani et al., 2018; L. Yu et al., 2014; Nery et al., 2016). Different ML algorithms have been used for urban LULC mapping and modeling (e.g. C. Zhang et al., 2019; Mao et al., 2020; Talukdar et al., 2020; Teluguntla et al., 2018), and have also been compared (Camargo et al., 2019; Li et al., 2016; Rogan et al., 2008). However, each ML algorithm will yield different accuracy levels for specific case studies and data. Further, in addition to the quality and quantity of the imagery, the choice of the suitable ML classifier is still a challenge as the hyperparameters of the classifiers influence the quality of the feature extractions (Lu & Weng, 2004; Nichols et al., 2019; Thanh Noi & Kappas, 2017).

To determine the suitable and most accurate ML approaches for urban LULC modeling, several studies have compared different classifiers based on their overall accuracy, and not in terms of their mathematical and functional approach (e.g. Camargo et al., 2019; Jamali, 2019; Li et al., 2016; Rogan et al., 2008). Comparing RF, SVM, naïve Bayes, and kNN machine learning classifiers for mapping urban areas in the city of Cape Town, Lefulebe et al. (2022) found all the classifiers to have accuracy of greater than 91%, with kNN being the best classifier at 96.54% accuracy with kappa of 0.95. Kranjčić et al. (2019) classified green infrastructure in Varaždin and Osijek cities in Croatia from Sentinel-2 satellite image using SVM, RF, ANN, and naïve classifier and found SVM to yield the highest accuracy of 87% with kappa coefficients of 0.89. Shi et al. (2019) used multisource satellite images for urban LULC mapping in Guangzhou by integrating an ensemble of object-based classifiers, decision trees and RF and achieved an accuracy of >85%. Ha et al. (2020) also used RF to map rural urbanization in Vietnam obtaining an accuracy of more than 90% using Landsat data. For detailed mapping of urban land use with multi-source data in the city of Lanzhou, Zong et al. (2020) used RF and attained overall accuracy of 83.75%. From previous research, RF, SVM and ANN have been reported to provide higher overall accuracy in urban LULC modeling as compared to the traditional classification techniques (Carranza-García et al., 2019; Gong et al., 2020; Ma et al., 2019).

In addition to their inherent mathematical functionalities, the performances of the classifiers are also influenced by the data and characteristics of the land-use features within the urban landscape. For example, simple decision trees-based classifiers like

Classification and Regression Trees (CART) are not only sensitive to changes in the training data sets but also tend to overfit the model (Prasad et al., 2006). On the other hand, for kNN classifiers, the setting of the ideal value of k is difficult (Naidoo et al., 2012), and the classifier is computationally complex as its effectiveness is dependent on the a priori determination of the number of neighbours (Qian et al., 2015). Naïve Bayes classifiers perform satisfactorily with small data sets; however, the output accuracy is compromised if inputs are not independent. SVM, on the other hand, is effective in high dimensional spaces and performs adequately in situations where a clear margin of separation exists between classes and is computationally efficient. Nevertheless, SVM requires a long training time for large data sets and is not intuitive, easy to understand or fine-tune (Huang et al., 2002). Though widely used due to their non-parametric modeling capability, ANN classifiers are highly complex, time-consuming and computationally expensive, require large training data to produce accurate outcomes and exhibit a high tolerance for noisy data (Y. Ouma et al., 2022).

Further, while several studies have been conducted on urban LULC mapping using ML algorithms, the performance of the models cannot be replicated from one case study to another, and most studies do not focus on classifier-class performance; rather, the emphasis is on the overall LULC classification accuracy. Secondly, most of the studies are based on single-date imagery and not on the multitemporal imagery with multisensor characteristics. Previous studies have also pointed out that the performance of the ML classifiers in urban LULC classifications are affected by the limitations in the spectral and spatial resolutions of the sensors especially at lower resolutions (Yang et al., 2017). Compared to the traditional classifiers, the non-parametric ML algorithms are considered superior as they do not rely on a priori hypotheses of the input data distribution (Nery et al., 2016). However, results from different case studies have demonstrated that the performance of a given ML classifier is not only specific to the case study, but also influenced by the setup of the machine learning algorithm itself and the quality of the training data. With focus on computationally efficient open-source solutions, this study evaluates the performance of Gradient Tree Boosting (GTB), RF, in comparison with SVM and Multilayer Perceptron Neural Networks (MLP-ANN) as implemented within the Google Earth Engine (GEE) platform for urban LULC mapping.

RF derives the optimal classification solutions by overcoming the limitations of single decision tree classifiers through robust ensemble learning (EL), majority voting and being able to handle higher number of model variables (Belgiu & Drăguț, 2016). GTB is

also an ensemble classifier, which, unlike RF, ensembles weak base learners with the aim of minimizing the loss function through adding newer weak learners to the ensemble (Friedman, 2002). GTB has the advantage of high-order feature information optimization, generalization and representation without scaling. Compared to other EL algorithms, for every GTB iteration, the negative gradient loss values are used to fit the residuals of the regression tree (Y. Ouma et al., 2022). SVM determines the best boundary between different training classes by features transfer to higher dimensions and performs well on high-dimensional data using dynamic kernel functions and are adaptable for different classification tasks (Pedregosa et al., 2011). Given the dense and complex nature of the input training data for urban LULC classification, MLP is considered to overcome potential classification overfitting scenarios (Mohtadifar et al., 2022). With focus on computationally efficient open-source solutions, this study evaluates the performances of the EL (GTB and RF) and non-EL (SVM and MLP) algorithms in mapping of urban LULC classes as implemented within the GEE platform for urban LULC mapping. Based on their performances for urban LULC feature mapping, the results of EL and non-EL algorithms will be evaluated for a proposed feature-in feature-out (FEI-FEO) post-classification fusion approach.

For the test study area of the Greater Gaborone Planning Area (GGPA) in Botswana, the objectives and contributions of this study are (1) to implement ensemble decision tree-based (GTB and RF classifiers) and SVM and MLP-ANN machine-learning classifiers for urban LULC mapping and change detection from Landsat data from 1984 to 2020 in 5-year intervals; (2) compare the performance of the classifiers in the

extraction of urban LULC classes at different multi-temporal scales and multisensor data; and (3) evaluate the significance of FEI-FEO post-classification fusion strategy in the extraction of urban features as optimally detected from the classifiers. This study improves on our previous study reported in Y. Ouma et al. (2022) by introducing GTB and ANN, and on the optimization of classifier hyperparameters for more accurate classification.

Materials and methods

Study area

The GGPA is the main urban area in Botswana with the highest population concentration (Figure 1). GGPA lies between latitude 20° 30'S and 24° 45'S and longitude 25° 50'E and 26° 12' (Figure 1), with an average altitude of 1,004 m AMSL. The GGPA land area is approximately 961.73 km², while the city is approximately 169 km². The spatial expansion of the city and the larger GGPA is constrained in part by the existence of the Gaborone dam, the traditional land tenure within the larger GGPA, and the topographic and semi-arid climate of the area. As depicted in Figure 1, within the commuting radius of the Gaborone city, a dormitory of suburbs are rapidly evolving which are resulting in the characteristic centripetal movement of rural –urban migrations.

Data

Multisensor Landsat series data comprising of Landsat-4 (L4-MSS), Landsat-5 (L5-TM), Landsat-7 (L7-ETM+) and Landsat-8 (L8-OLI) acquired from 1984 to 2020

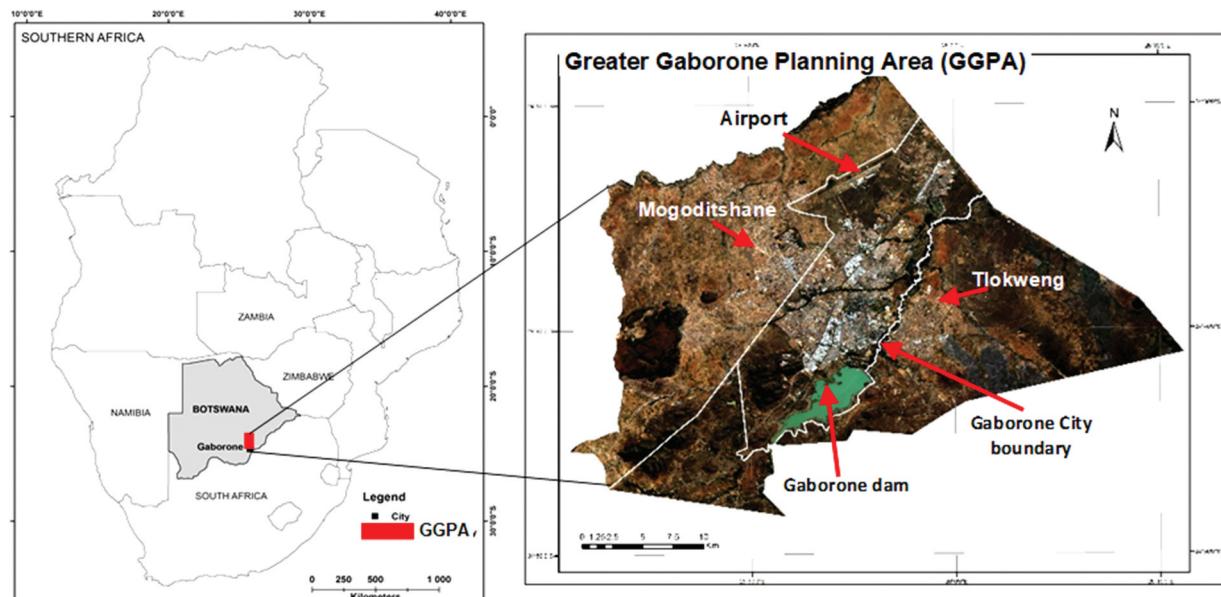


Figure 1. Location of the test study area: greater gaborone planning area (GGPA) (reprinted with permission from Y. Ouma et al. (2022). Copyright 2022 ISPRS archives).

Table 1. Characteristics of Landsat MSS, TM, ETM+ and Landsat OLI sensors.

Landsat sensors	Spectral wavelength (μm)	Spectral bands	Spatial resolution (m)	Temporal period (5-year)
Landsat MSS	0.50–0.60	Green	60	1984
	0.60–0.70	Red		
	0.70–0.80	Near IR1		
	0.80–1.11	Near IR2		
Landsat TM	0.45–0.52	Blue	30	1990, 1995, 2000, 2005
	0.53–0.60	Green		
	0.63–0.69	Red		
	0.75–0.90	Near IR1		
	1.55–1.75	SWIR1		
	2.09–2.35	SWIR2		
Landsat ETM+	0.45–0.52	Blue	30	2010
	0.52–0.60	Green		
	0.63–0.69	Red		
	0.77–0.90	Near IR1		
	1.55–1.75	SWIR1		
	2.09–2.35	SWIR2		
Landsat OLI	0.45–0.51	Blue	30	2015, 2020
	0.53–0.59	Green		
	0.64–0.67	Red		
	0.85–0.88	Near IR2		
	1.57–1.65	SWIR1		
	2.11–2.29	SWIR2		

were downloaded from the USGS Earth explorer. The MSS pixel size of 60 m was resampled into 30 m using data fusion as implemented in Chen et al. (2017). To minimize the seasonality effects, the data sets were acquired during the time of year. The multitemporal and multisensor Landsat imagery (Table 1) were atmospherically corrected using the ATCOR2 tool and histogram equalization in ERDAS Imagine. The time-series bands were mosaiced, composited, resampled to 30 m spatial resolution and clipped to the study area.

Training and testing data

The urban LULC classes comprised of built-up (residential, commercial, industrial and impervious surfaces); water (dam water body), vegetation cover (forest, shrubs and grass) and bare soil. Depending on the season, the grass and shrubs are partially converted to croplands. The aggregation of the built-up components was adopted to

minimize the spectral mixing and spatial heterogeneity of the built-up features. It was, however, possible to spectrally distinguish between the vegetation types due to their expansive areal extents and the resulting spectral homogeneity. Figure 2 presents the LULC classes and the spectral reflectance trends for the classes in the Landsat's visible, NIR and SWIR bands. For each year, the training comprised of 12,500 pixels and 7,500 pixels were used for model accuracy testing. The training and testing data samples were collected from visual identification and interpretations from the Landsat imagery, Google Earth historical imagery and the historical LULC maps.

Methods

Decision tree-based machine learning classifiers

This section presents an overview of the functional and implementational approaches for GTB, RF, SVM and MLP-ANN classifiers. To improve on the

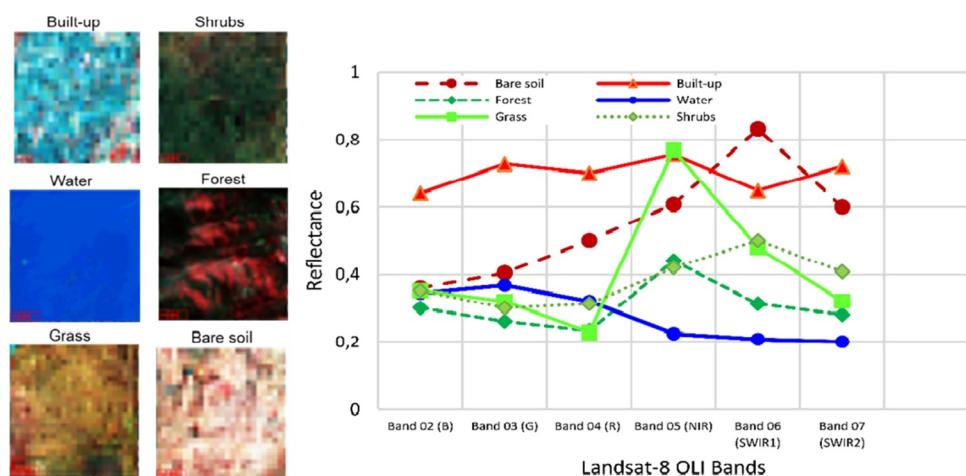


Figure 2. False color composite LULC classes and the spectral reflectances in L8-OLI (part of figure reprinted with permission from Y. Ouma et al. (2022). Copyright 2022 ISPRS archives).

performance accuracy of the classifiers, optimal model hyperparameters determined following the steps outlined in Y. O. Ouma et al. (2023).

Gradient tree boosting (GTB)

GTB aggregates an ensemble of decision trees (Figure 3). The classifier, however, confines individual trees to a weaker prediction model, hence limiting the complexity of the decision trees. The algorithm attains its classification accuracy by the iterative combination of weak learner ensembles into stronger ensemble of trees through stepwise minimization of the loss function based on the gradient descent optimization (Friedman, 2002). Different from the other ensemble classifiers, GTB fits residuals of the regression tree at each iteration using negative gradient values of loss. The inter-tree correlations are reduced by constructing new trees based on stochastically selected training subset data.

The GB algorithm for training and classification of sample feature vector is summarized in the following steps:

- (I) Determine the training T and testing S sets, i.e. $T = (x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)$, $x_i \in X \subseteq R^n$, $y_i \in \{0, +1\}$. Let x_i denote the feature vector of each sample and y_i denotes its class label.

- (II) Establish the loss function:

$$L(y, f(x)) = [y - f(x)]^2 \quad (1)$$

where $f(x)$ is the fitting function of y .

III. Initialization of the model variables:

$$f_o(x) = \arg \min_c \sum_{i=1}^N L(y_i, c) \quad (2)$$

where c is the constant for the minimization of the loss function $L(y, c)$ and represents a tree with a single root node.

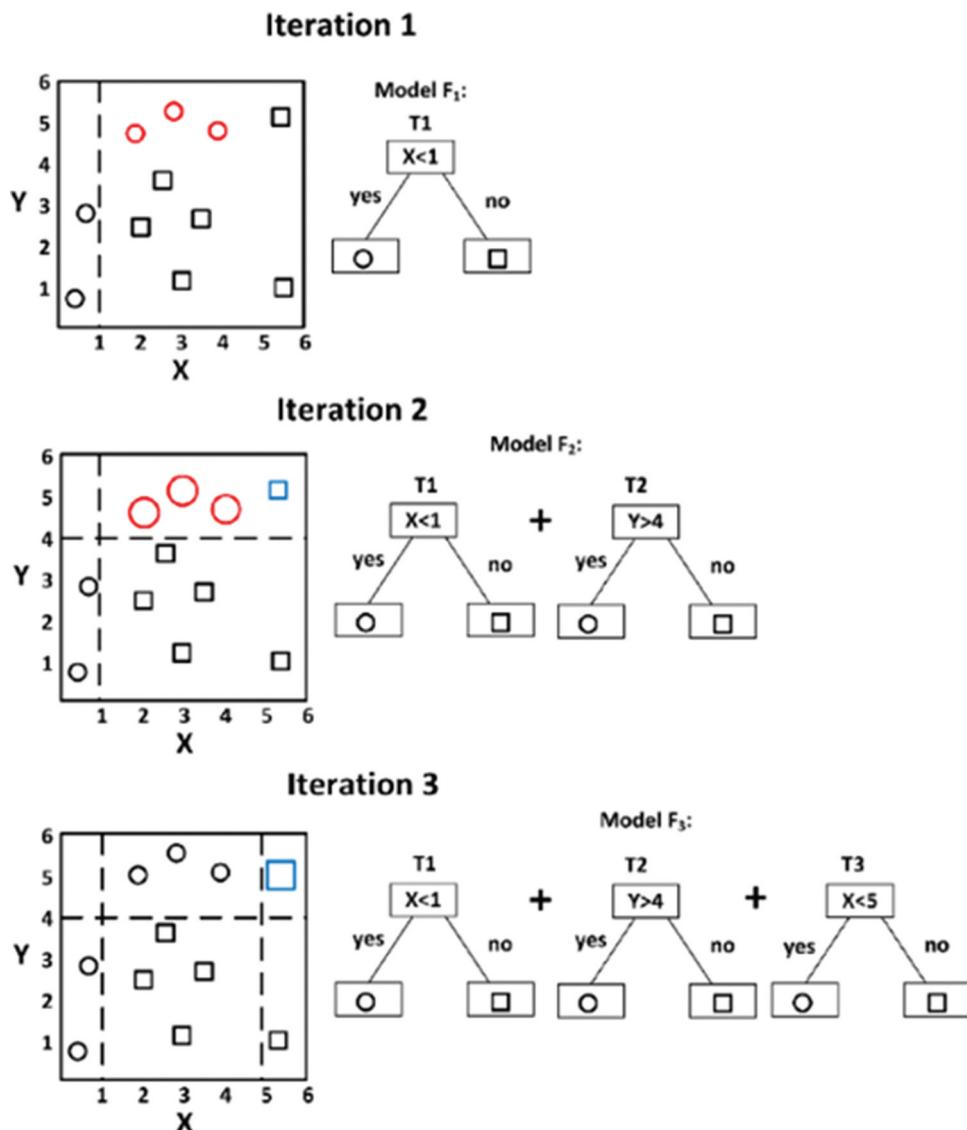


Figure 3. Visualizing gradient tree decision boosting (reprinted with permission from Y. Ouma et al. (2022). Copyright 2022 ISPRS archives).

IV. For each model the following steps are executed:

(a) For the i -th sample ($I = 1, 2, \dots, N$), calculate the negative gradient r_{mi} of the loss function in the current model r_{mi} :

$$r_{mi} = - \left[\frac{\partial L(y_i, f(x_i))}{\partial f(x_i)} \right]_{f(x)=f_{m-1}(x)} \quad (3)$$

where m is the model number, with $m = 1, 2, \dots, M$, M is the maximum value of m and J is the maximum value of j .

- (a) Fit a regression tree for r_{mi} and determine the leaf node area r_{mj} ($j = 1, 2, \dots, J$) for the m^{th} tree.
- (b) Number of leaf nodes j ($j = 1, 2, \dots, J$) is determined from Eq. 4:

$$c_{mj} = \arg \min_c \sum_{x_i \in R_{mj}} L(y_i, f_{m-1}(x_i) + c) \quad (4)$$

(d) To minimize the loss function and estimate the value of the leaf node area, a linear search is used in this step.

Tree update according to Eq. (5):

$$f_m(x) = f_{m-1}(x) + \sum_{j=1}^J c_{mj} I(x \in R_{mj}) \quad (5)$$

where, if $x \in R_{mj}$, then $I = 0$ or $I = 1$.

V. Determination of the final tree model (Eq. (6)):

$$\hat{f}(x) = f_M(x) = \sum_{m=1}^M \sum_{j=1}^J c_{mj} I(x \in R_{mj}) \quad (6)$$

From an ensemble of weaker models, GB creates new models such that each of the created models minimizes the loss function $L(y, f(x))$, to fit a more accurate model with improved overall accuracy. To minimize overfitting, a boosting threshold criterion based on either the achieved prediction accuracy or the maximum number models to be created is adopted. The weak learning process of GTB decision tree is complemented by improving the representation, optimization and generalization so as to capture the higher-order information and is invariant to scaling of sample data. Further, by weighting the combination scheme, GTB can avoid overfitting by fitting the residuals of the regression tree at each iteration using the negative gradient values of loss.

Random forest (RF)

RF is made up of combined multiple decision trees (Figure 4) trained upon random subsets of the labeled samples and features (Breiman, 2001). A decision tree itself is a deterministic data structure for modeling

decisions rules. At each internal node of the decision tree, a feature is chosen to infer the class determining the specific decision by splitting the incoming training samples to maximize the information gain. Similarly, each tree in the ensemble is constructed from a sample (bootstrap sample) that is replaceably drawn from the training set. The RF training process is similar to CART; however, to increase computational efficiency in RF, each tree only utilizes a random subset of features at each node to reduce correlation (Figure 4).

Let $u \in U \subset \mathbb{R}^n$ be an input feature vector, and $v \in V \subset \mathbb{R}$ be its corresponding target value for regression. For a given internal node j and a set of samples $S_j \subset U \times V$, the information gain achieved by choosing the k th feature to split the samples in the regression problem is computed according to Eqs. (7–9):

$$I_j^k = H(S_j) - \frac{|S_{j,L}^k|}{|S_j|} H(S_{j,L}^k) - \frac{|S_{j,R}^k|}{|S_j|} H(S_{j,R}^k) \quad (7)$$

$$H(S) = \frac{1}{|V|} \sum_v (v - \bar{v})^2 \quad (8)$$

$$\bar{v} = \frac{1}{|V|} \sum_v (v) \quad (9)$$

where L and R denote the left and right child nodes, $S_{j,L}^k = \{(u, v) \in S_j \mid u^k < \theta_j^k\}$, $S_{j,R}^k = S_j \setminus S_{j,L}^k$, u^k is the k th feature of the feature vector u , θ_j^k is the splitting threshold chosen to maximize the information gain I_j^k for the k th feature u^k , and $|\cdot|$ is the cardinality of the set. $H(S)$ denotes the variance of all target values in the classification or regression problem.

In training the RF algorithm, the splitting θ_j^k is implemented recursively until either the information gain I_j^k is insignificant or the training samples input into a given node is less than the preceding threshold θ_{j-1}^k . Using the out-of-bag errors is often adopted as the option for parameter tuning in (Biau & Scornet, 2016). The advantage of RF is that it can produce stable, robust and accurate results even with minimal tuning of the hyperparameters. The algorithm is easy to parameterize, insensitive to over-fitting and deals with outliers in training data, reporting the classification error and variable significance. Further, RF is able to process multidimensional features from both continuous and categorical datasets. In implementing RF, predictions for classification are performed by obtaining and bagging the majority class vote from the individual tree class votes. The output classification result for a new sample is obtained through a majority voting of the individual tree results. For RF tuning, the

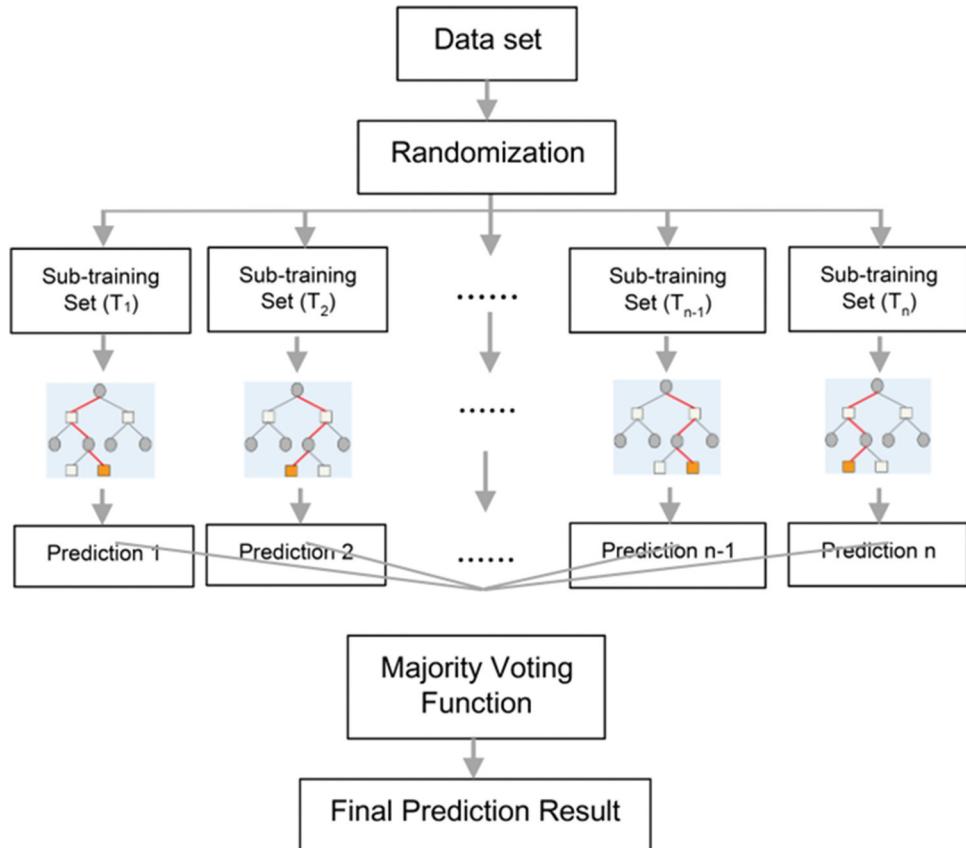


Figure 4. Illustration of the random forest classification structure.

number of trees (n_{Tree}) and the number of variables per split (m_{Try}) are optimized.

Support vector machine (SVM)

The SVM model fits an optimal separating hyperplane or hyperplanes in a high-dimensional space, to create an optimal boundary between two classes that enables the prediction of labels from one or more feature vectors (Noble, 2006). The hyperplane(s) are orientated furthest from the closest data points from each of the classes. These closest points are the support vectors. Like the ensemble algorithms, SVM comprises of a set of related learning algorithms for classification and regression. Given a labeled training data set:

$$T = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}, x_i \in R^d \text{ and } y_i \in \{-1, +1\} \quad (10)$$

where x_i is a feature vector representation, y_i the class label of a training compound I and n is the elements in the training data sets. The optimal hyperplane is defined by

$$wx^T + b = 0 \quad (11)$$

where w is the weight vector, x is the input feature vector and b is the bias. w and b , respectively, satisfy

the inequalities for all elements of the training set as (Eqs. 12–13).

$$wx_i^T + b \geq +1, \text{ if } y_i = +1 \quad (12)$$

$$wx_i^T + b \geq -1, \text{ if } y_i = -1 \quad (13)$$

The aim of training in SVM model is to determine the w and b so that the hyperplane separates the data and maximizes the margin $1/\|w\|^2$. Vectors x_i for which $(wx_i^T + b) = 1$ will be termed support vector as depicted in Figure 5.

By solving the optimization task (Eq. 14), the linear SVM determines the optimal separating margin:

$$\begin{aligned} & \text{minimize} \left\{ \frac{1}{2} |w|^2 + C \sum_{i=1}^n \varepsilon_i \right\}, \varepsilon_i \\ & \geq 0 \text{ subject to } y_i (\mathbf{w}^T x_i + \mathbf{b}) \geq 1 - \varepsilon_i, i \\ & = 1, 2, \dots, n \end{aligned} \quad (14)$$

where C is the optimum cost parameter, ε_i defines the positive slack variables, \mathbf{w} is a normal vector and \mathbf{b} is a scalar quantity. For n_{SV} support vectors, \mathbf{W} becomes a linear combination of the training vectors (Eq. 15), and \mathbf{b} the average of all support vectors (Eq. 16):

$$\mathbf{W} = \sum_{i=1}^n \alpha_i y_i x_i \quad (15)$$

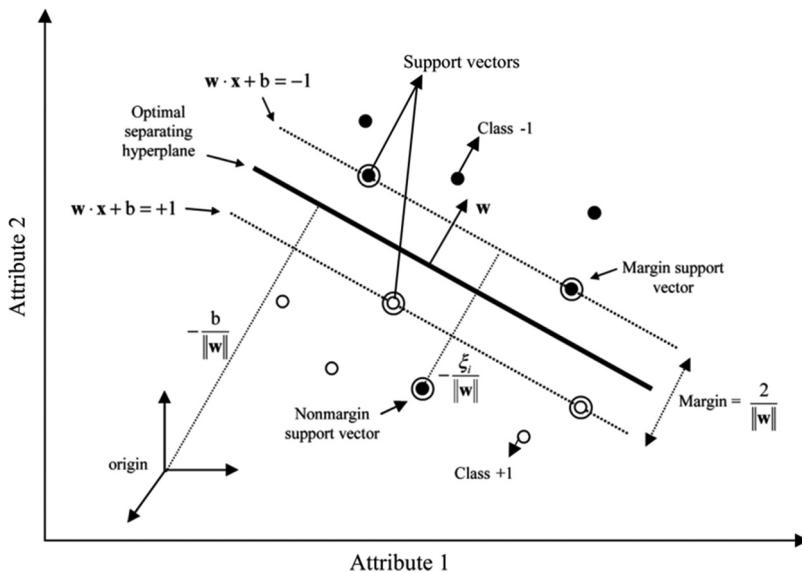


Figure 5. Maximum margin-minimum norm classifier in support vector machine with optimal hyperplane for linearly non-separable classes (reprinted with permission from Y. Ouma et al. (2022). Copyright 2022 ISPRS archives).

$$\mathbf{b} = \frac{1}{N_{SV}} \sum_{i=1}^{N_{SV}} (Wx_i - y_i) \quad (16)$$

To expand the conventional linear SVM into the nonlinear cases, x_i is replaced through mapping into the feature space $\theta(x_i)$, such that $x_i^T x$ is expressed in the form $\theta(x_i)^T \theta(x_i)$ in the transformation feature space. The nonlinear discriminate function is expressed as in Eq. 17, where $K(x_i, x) = \langle \theta(x_i), \theta(x) \rangle$ and $K(x_i, x)$ defines the kernel function.

$$f(x) = \text{sgn}\left(\sum_{i=1}^n \alpha_i y_i K(x_i, x) + \mathbf{b}\right) \quad (17)$$

For better performance of SVM classifier in land-cover classification, Knorn et al., 2009 demonstrated that RBF kernel function is preferred due to accuracy and reliability (X. Yu et al., 2004), and was adopted in the current study. The $K(x_i, x)$ is defined as

$$K(x, y) = \langle f(x), f(y) \rangle \quad (18)$$

where K is the kernel function, x, y are n -dimensional inputs; f is used to map the input from the n -dimension to m -dimensional space, and $\langle x, y \rangle$ denotes the dot product.

The kernel functions are used to calculate the scalar product between two data points in a higher-dimensional space without explicitly calculating the transformation from the input space to the higher dimensional space. The kernel computation is easier in the high dimensional space for the determination of the inner product of two feature vectors. This is an advantage in the complexity in computing the feature vectors for kernels. For the RBF kernel $K_{RBF}(x, y) = \exp(-\gamma \|x - y\|^2)$, though the corresponding feature vector is infinite dimensional, the kernel computation is trivial.

In implementing the SVM classifier with the RBF kernel, there are two main determinants (Ballanti et al., 2016; Qian et al., 2015):

- (i) Optimum cost parameter C which determines the size of the allowed misclassification for spectrally overlapping training data to enable the possible adjustment of the training data. To minimize model over-fitting larger C values are preferred.
- (ii) Kernel width parameter γ determines the degree of smoothing and shape of the hyperplane dividing the class (Melgani & Bruzzone, 2004). Increasing the γ affects the shape of the class-dividing hyperplane which may influence the accuracy of the classification results.

Multilayer perception neural network (MLP-ANN)

Artificial neural networks (ANN) have different topological structures including multilayer perceptron (MLP), adaptive neuro fuzzy inference system (ANFIS), generalized regression neural networks (GRNN), recurrent neural networks (RNN), and radial basis function network (RBFN). These ANN models can generally be categorized into feedforward neural networks (FFNN) and RNN. The most popular FFNN is the MLP-ANN trained with a backpropagation learning algorithm. FFNN have the advantages that with single or few hidden layers suitable activation functions, the model can approximate a complex and nonlinear system. The adopted MLP-ANN model for this study is represented in Figure 6.

In Figure 6, R = total number of inputs; z = hidden neurons; $\omega_{i,j(1)}$ = weight of first layer between the input j and the i th hidden neuron; $\omega_{i,j(2)}$ = weight

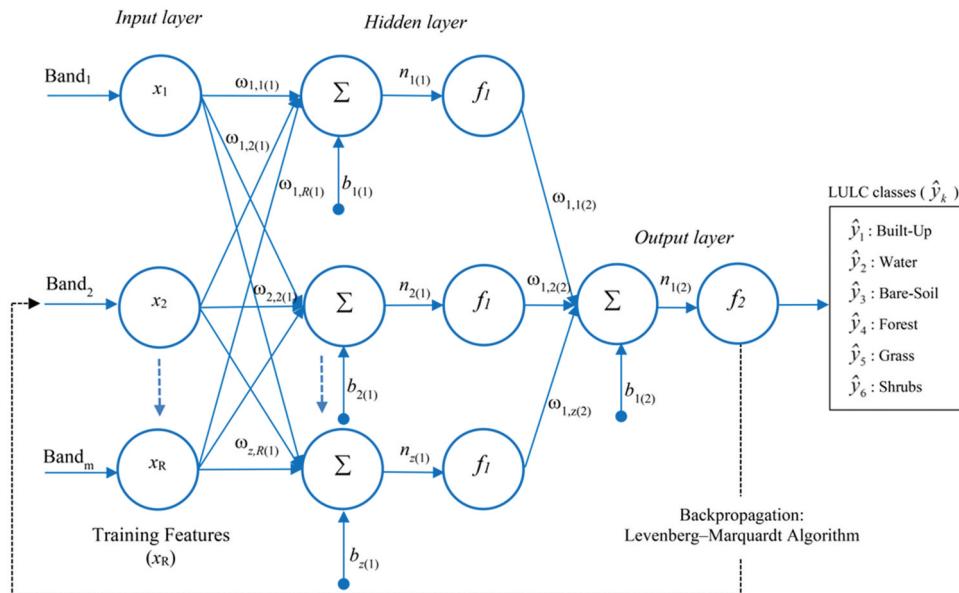


Figure 6. Topology of a three-layer MLP-ANN.

of second layer between the i th hidden neuron and output neuron; $b_{i(1)}$ =bias weight for the i th hidden neuron; and $b_{1(2)}$ =bias weight for the output neuron.

Multi-classifier and multi-feature fusion approach

This study proposes a multi-classifier and multi-feature fusion based on the concept of feature in-feature out (FEI-FEO) post-classification feature fusion approach where the extracted features with the highest accuracy are combined under mutual exclusivity conditions. The rationale behind FEI-FEO is that when the scene information from different classifiers $\{C_1(X), C_2(X), \dots, C_M(X)\}$ represents different parts of the scene at different accuracies, the features extracted with the highest accuracies can be combined to obtain the complete and more accurate global information for a given LULC scene. The proposed FEI-FEO based fusion approach relies on the complementary combination of the input features $\{F_{m1}, F_{m2}, F_{m3}, \dots, F_{mn}\}$ to produce the new scene global features $(F_1, F_2, F_3, \dots, F_n)$ with the highest extraction accuracies. The FEI-FEO feature fusion approach is empirically illustrated in Figure 7.

The FEI-FEO data fusion process addresses a set of features with the aim to improve, refine or obtain new features (Dasarathy, 1997). The advantage of the optimal FEI-FEO fusion approach is on the fact that the optimal class is detected, and this can permit generalization of the classifier for class or feature detection with even lesser number of training samples.

Accuracy assessments

For evaluation of the performance of the classifiers, confusion matrices were generated based on

a crosscheck between the classification results and test samples. The following metrics are used: (i) producer's accuracy (PA); (ii) user's accuracies (UA); (iii) overall accuracy (OA); (iv) kappa index; and (v) F1-score (Eqs. 19–22). PA measuring the degree of precision is the proportion of the samples that truly belong to a specific class among all those classified as that specific class, while UA or recall is the proportion of samples classified as a specific class among all the samples that truly belong to that class (Nevalainen et al., 2017). The OA is the number of correctly classified samples to the total number of samples. The kappa index provides the agreement of prediction with the true class, considering the random chance of correct classification, and the F1-measure is the harmonic mean of precision and recall and was calculated to determine the performance at a classifier and class levels.

$$UA = \frac{K_{ii}}{K_{i+}}, \text{ and } PA = \frac{K_{jj}}{K_{j+}} \quad (19)$$

$$OA = \frac{\sum_{i=1}^n K_{ii}}{T} \quad (20)$$

$$K = \frac{T \sum_{i=1}^n K_{ii} - \sum_{i,j=1}^n (K_{i+} K_{+j})}{T^2 - \sum_{i,j=1}^n (K_{i+} K_{+j})} \quad (21)$$

$$F1 - score = 2 * \frac{UA * PA}{UA + PA} \quad (22)$$

where PA = producer's accuracy; UA = user's accuracy; OA = overall accuracy; K = Kappa coefficient; n = number of classifications; K_{ii} = number of correct classification; K_{i+} = number of pixels in the i th row and K_{+j} = number of pixels in the j th column; and T = number of pixels used for the accuracy evaluation.

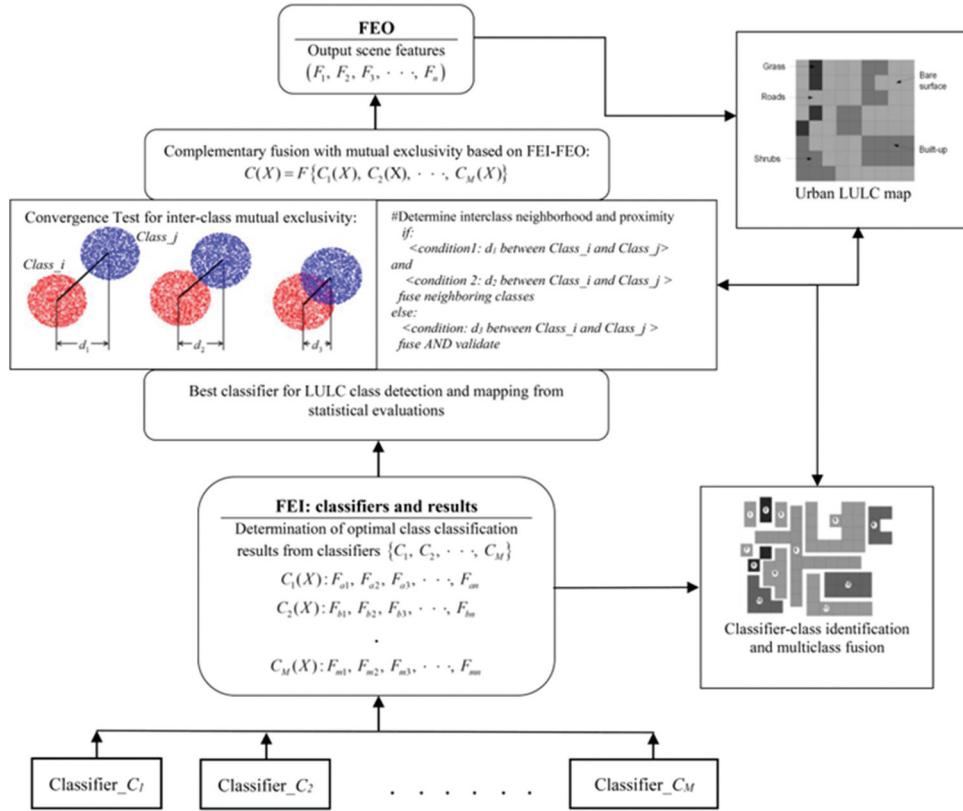


Figure 7. The feature in-feature out (FEI-FEO) fusion for post-classification fusion.

The statistical significance of the differences in the classification accuracy between the classifiers was evaluated using pairwise z -score test. The z -test was applied to the OA results for testing statistical significance at a significance level of 5%. If $z > 1.96$, the test is significant, leading to the conclusion that the obtained results differ from each other.

Results

Influence of training data size on classifier performance

In this section, the accuracy for urban LULC mapping is considered as a function of the number of training

samples or batch size required for reliable labeling, training and classification. The optimization of the training samples is important as it is time-consuming, costly to obtain requires more storage and computation power. Varying the training size, the classifier accuracy outputs were averaged with the results in presented in Figure 8.

The classification accuracy is observed to improve for all the classifiers as the number of training samples increases and tended to converge to an optimal classification rate when the number of training samples was above 10,500 pixels, with all the classifiers performing at >80% in overall accuracy. The satisfactory performance with higher number of training samples is attributed to the

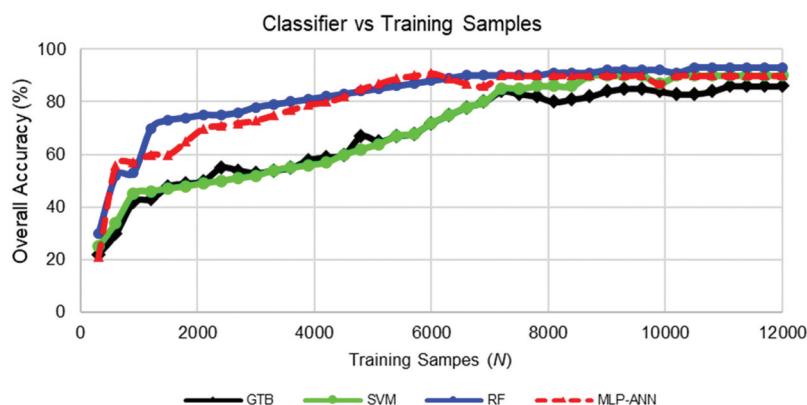


Figure 8. Variation of number of training samples and classifier overall average accuracy.

process required to minimize the separability, noise and time invariance of the sound features. The results in Figure 8 depict the significance of training samples as a critical hyperparameter in the implementation and comparison of machine learning classifiers.

Accuracy assessment

Class producer and user accuracies

The results for the PA and UA for each class in the respective years are presented in Figure 9. For

detecting built-up areas using the PA metrics, MLP-ANN had the highest average of PA 93.2%, followed by RF (90.6%), SVM (89.3%) and GTB (85%). In terms of the UA, MLP-ANN and RF had the highest values of 93% and 92.3%, respectively, while SVM and GTB had respective UAs of 91.9% and 86.1%. The water bodies were classified with consistently higher average PA accuracy of 99% by MLP-ANN, followed by RF (98.1%), GTB (97.5%) and SVM (97.2%) and the corresponding UA values were RF and MLP-ANN equally at 99.7%, SVM (99.3%) and GTB (95.3%).

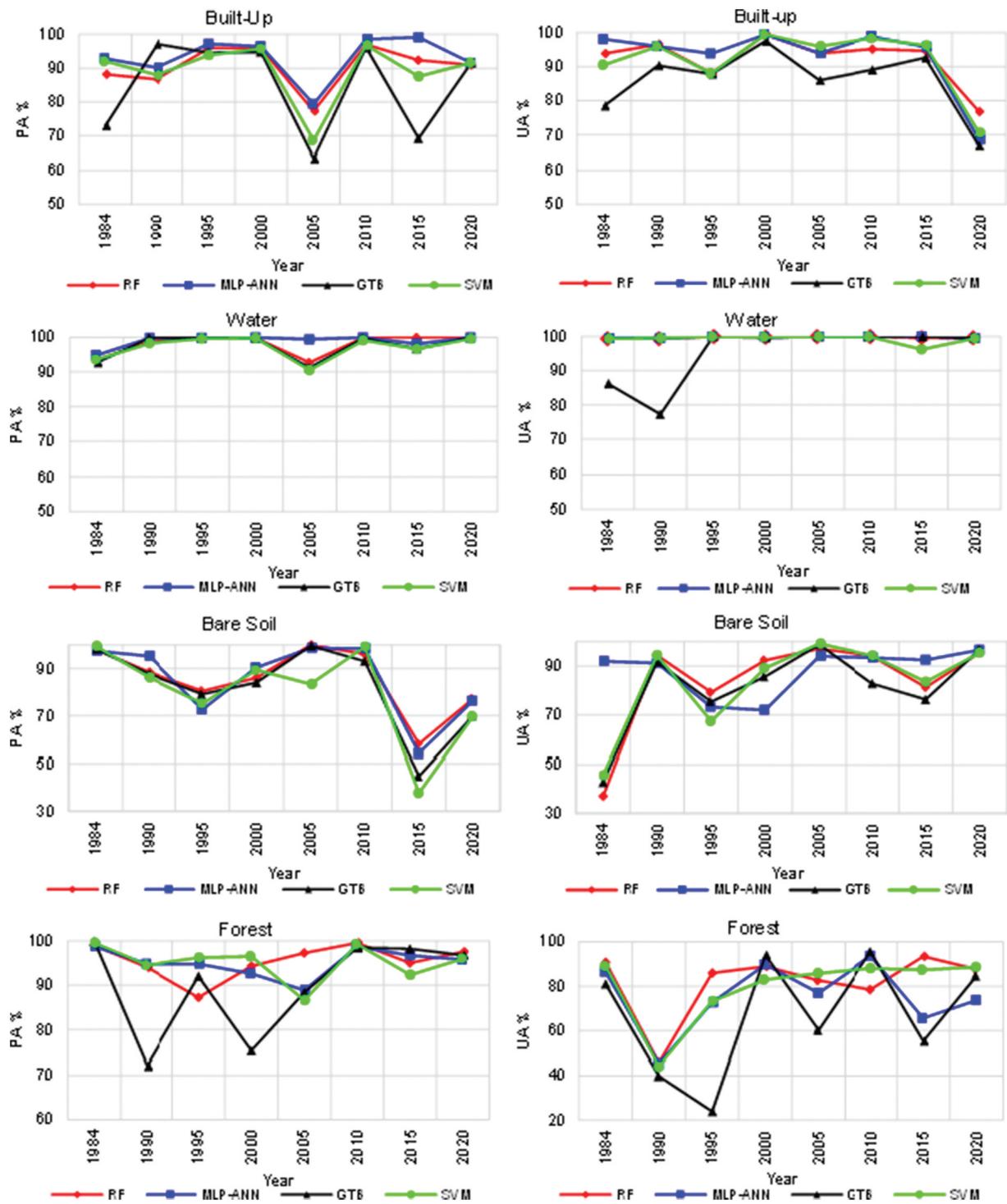


Figure 9. Average yearly PA and UA metrics for LULC classes.

For bare-soil mapping, RF attained average PA of 85.6% which was 0.1%, 3.5% and 5.6% higher than MLP-ANN, GTB and SVM, respectively. The UA for bare soil was highest for MLP-ANN (88.1%) and followed by RF (83.9%), SVM (83.7%) and GTB (81.3%). In mapping the bare-soil cover, all the classifiers achieved lower PAs compared to built-up areas and waterbodies classes. This could be attributed to the spectral confusion of bare-soil with the impervious surfaces including buildings and roads. For the vegetation classes, all the classifiers attained the lowest average PA of 83.5%, with grass having the highest average PA of 88% and shrubs having the least PA of 80.5%. In terms of the UA, the average accuracy was 85.3% and the trend was the same with grass (92.3%), forest (83.2%) and shrubs (80.3%). The low PA and UA in vegetation mapping is contributed to by the intra-spectral confusion in the vegetation classes.

Apart from the spectrally homogeneous water body, the PA and UA measures are observed to vary with the urban LULC class, time, sensor and the classifier. This requires further statistical evaluation of the classification results to determine the suitability of the classifiers for detecting and extracting specific urban LULC classes.

Average class classification accuracy

Table 2 presents the average metrics in terms of OA, F1-score, TPR (true positive rate or sensitivity), FPR

(false positive rate) and AUC (area under the roC curve) measures for each class for the 8-epochs. All the classifiers mapped water bodies with the highest OA, F1-score, TPR and AUC scores for the 35-year period. For the built-up class, MLP-ANN had the highest accuracy. The vegetation classes were mapped with the least accuracy as compared to the other classes, with MLP-ANN being the best classifier for mapping grass, and RF being the most suitable for mapping shrubs and forest. MLP-ANN is recorded to be best classifier for detecting bare soil at 98.1%, which is 1.1% higher than the least accuracy from SVM.

For the overall average mapping of the LULC classes, RF achieved the highest performance with OA of 92.8%, MLP-ANN (91.2%), SVM (90.9%) and GTB (87.8%). The same performance pattern was observed from the overall average F1-score, TPR and AUC except for the FPR where GTB tended to have lower FPR values compared to the better performing classifiers per LULC class. The results of the TPR and FPR are presented in Figure 10 in terms of the area under ROC for all the classification models. On average for all the classes, the RF model had the highest area under ROC curve of 0.981, which is 0.021, 0.029 and 0.077, respectively. higher than MLP-ANN, SVM and GTB classifiers in mapping the urban LULC classes. The results in Figure 10 are the average results after tuning the classifier hyperparameters to yield the best results for a given year and for the urban LULC classes.

Figure 11 presents the summary of the F1-scores for each classifier, class and year. RF had the highest

Table 2. Classifier-class average accuracy.

Built-up	OA (%)	F1-score	TPR	FPR	AUC
GTB	97.2	0.848	0.861	0.018	0.921
RF	98.6	0.912	0.923	0.008	0.957
MLP-ANN	99.0	0.928	0.931	0.004	0.963
SVM	98.5	0.901	0.919	0.009	0.955
Water					
GTB	99.0	0.962	0.960	0.953	0.974
RF	99.6	0.989	0.970	0.997	0.997
MLP-ANN	99.8	0.994	0.995	0.997	0.998
SVM	99.5	0.984	0.987	0.997	0.996
Grass					
GTB	97.2	0.633	0.604	0.008	0.798
RF	97.8	0.811	0.826	0.010	0.908
MLP-ANN	97.9	0.772	0.755	0.006	0.874
SVM	97.0	0.720	0.728	0.010	0.859
Shrubs					
GTB	90.6	0.817	0.920	0.146	0.887
RF	94.7	0.867	0.939	0.076	0.931
MLP-ANN	92.8	0.844	0.941	0.102	0.920
SVM	93.9	0.857	0.921	0.079	0.921
Forest					
GTB	91.7	0.740	0.668	0.020	0.824
RF	95.4	0.872	0.816	0.010	0.903
MLP-ANN	94.2	0.833	0.754	0.012	0.871
SVM	95.0	0.860	0.799	0.013	0.893
Bare-soil					
GTB	97.2	0.795	0.813	0.017	0.898
RF	97.9	0.826	0.839	0.012	0.914
MLP-ANN	98.1	0.858	0.881	0.014	0.933
SVM	97.0	0.791	0.837	0.022	0.907

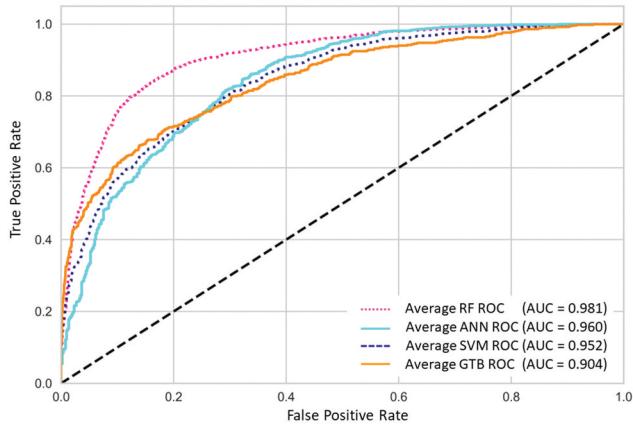


Figure 10. Average ROC curves for the classifiers.

average F1-score of 0.879 performing marginally higher than MLP-ANN by 0.007. SVM and GTB, respectively, recorded average F1-scores of 0.852 and 0.799. For the specific years, the performance of the classifiers varied with MLP-ANN recording the highest F1-score of 0.994 and the least F1-score of 0.633 was from GTB and GTB also recorded the least F1-scores for all the years. In overall from the F1-score measures, water was the best mapped land-cover for all the years. Based on the average class accuracy measures in Table 2, the best classifier for detecting a specific urban LULC class can be identified for a given case study.

Overall accuracy and kappa index

The overall accuracy results are presented in Figure 12, indicating that for all the years and

classes, RF performed better than all the classifiers with average OA of 92.8%. It is, however, observed that only MLP-ANN marginally outperformed RF in 1984, 1990 and 2010. In terms of the OA, RF performed better than MLP-ANN by 1.6%. In close performance to MLP-ANN is SVM with average OA of 90.9% and GTB had average OA of 87.8% performing lower than the other classifiers by between 3 and 5%. The kappa indices in Table 3 show that RF and MLP-ANN had average kappa index values of 0.880 and 0.863, while SVM and GTB achieved kappa indices of 0.860 and 0.794, respectively. Notable is the similarity in the performance trend between the classifiers with the varying multitemporal scales and multisensor data sets.

The results for the inter-comparison of the ML classifiers using the pairwise z-score test, such

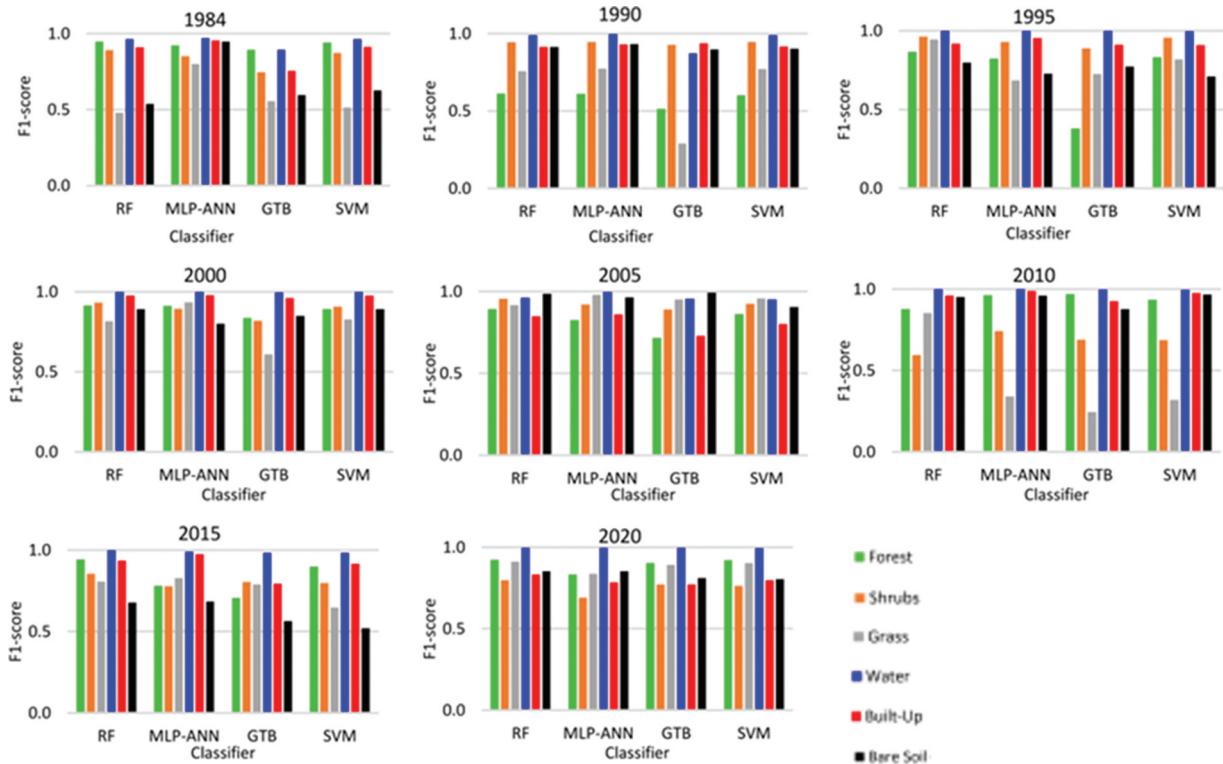


Figure 11. Classifier-class-year F1-scores.

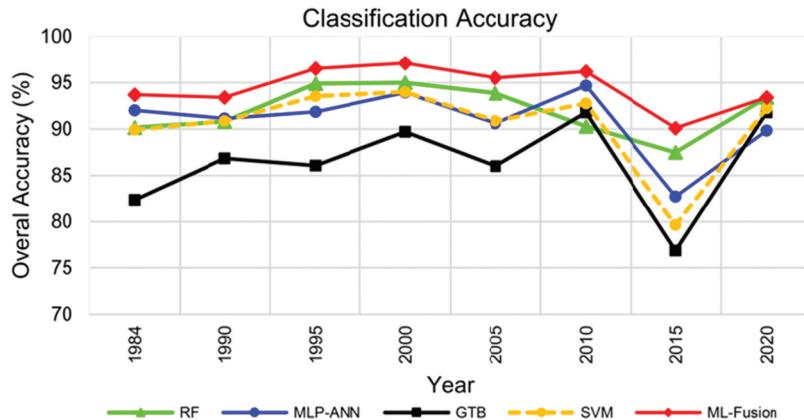


Figure 12. Average overall accuracy performance of the classifiers and the FEI-FEO fusion.

Table 3. Average kappa coefficients for the classifiers per year.

Year	Kappa index			
	GTB	RF	SVM	MLP-ANN
1984	0.761	0.864	0.862	0.891
1990	0.696	0.799	0.801	0.806
1995	0.767	0.921	0.898	0.872
2000	0.868	0.936	0.923	0.922
2005	0.768	0.900	0.855	0.846
2010	0.893	0.875	0.905	0.930
2015	0.717	0.845	0.751	0.788
2020	0.884	0.907	0.890	0.856

that $z\text{-score} > 1.96$ is considered statistically different at $\alpha = 0.05$ level of significance, are presented in **Table 4** with the average results that for the case study, there was no significant difference between the classifiers in terms of the overall accuracy of performance. The notable significant difference is between GTB and the other classifiers with a $z\text{-score} > 1$, and the least difference is between RF and MLP-ANN at $p\text{-value} = 0.881$.

LULC classification and change detection results

The urban LULC mapping and change detection results for the GGPA are presented in **Figure 13**, with the area coverage and percent change in area as mapped by each classifier. The spatial-temporal trends for each class are briefly discussed below.

Urban built-up: In 1984, RF and SVM estimated the built-up area to be 2.4% of the total area ($\sim 22.6 \text{ km}^2$), while both GTB and MLP-ANN were at 2.6% (25.3 km^2) as shown in **Figure 13**. All the classifiers showed progressive growth in urban area development in the successive years with the highest growth change recorded between 1984–1990 ranging from 62 to

72% for all the classifiers. The classifiers reported different statistics for the least growth period with RF and MLP-ANN showing that between 2005 and 2010 the urban growth was at 4% and 6%, respectively, while during the same period GTB and SVM showed double digit growth at 15% and 13%, respectively (**Figure 13**). In latest year 2020, the total built-up area was mapped by the classifiers as MLP-ANN (108.14 km^2), RF (113.39 km^2), GTB (135.17 km^2) and SVM (126.23 km^2). Given that RF detected the urban areas with the highest accuracy in 2020, it implies the other classifiers overestimated the built-up area. Though the built-up class mixing increases the classification accuracy, it does not capture the specific elements of the urban built-up ecosystem which comprises of residential, commercial, industrial and impervious surfaces.

Water: The main water body within the study area is the Gaborone dam. For all the classifiers, the homogeneous water body was mapped with highest accuracy 99% on average. In terms of the changes in surface area, there are observed fluctuations in the dam area and the results show that all the classifiers recorded maximum increase in the surface area in the

Table 4. Average z-scores and p-values for ML model pairs.

Classifier pairs	z-score	p-value	Significance level
GTB-RF	-1.687	0.092	No
GTB-SVM	-1.308	0.191	No
GTB-ANN	-1.476	0.140	No
RF-SVM	0.362	0.717	No
RF-ANN	0.150	0.881	No
SVM-ANN	-0.194	0.846	No

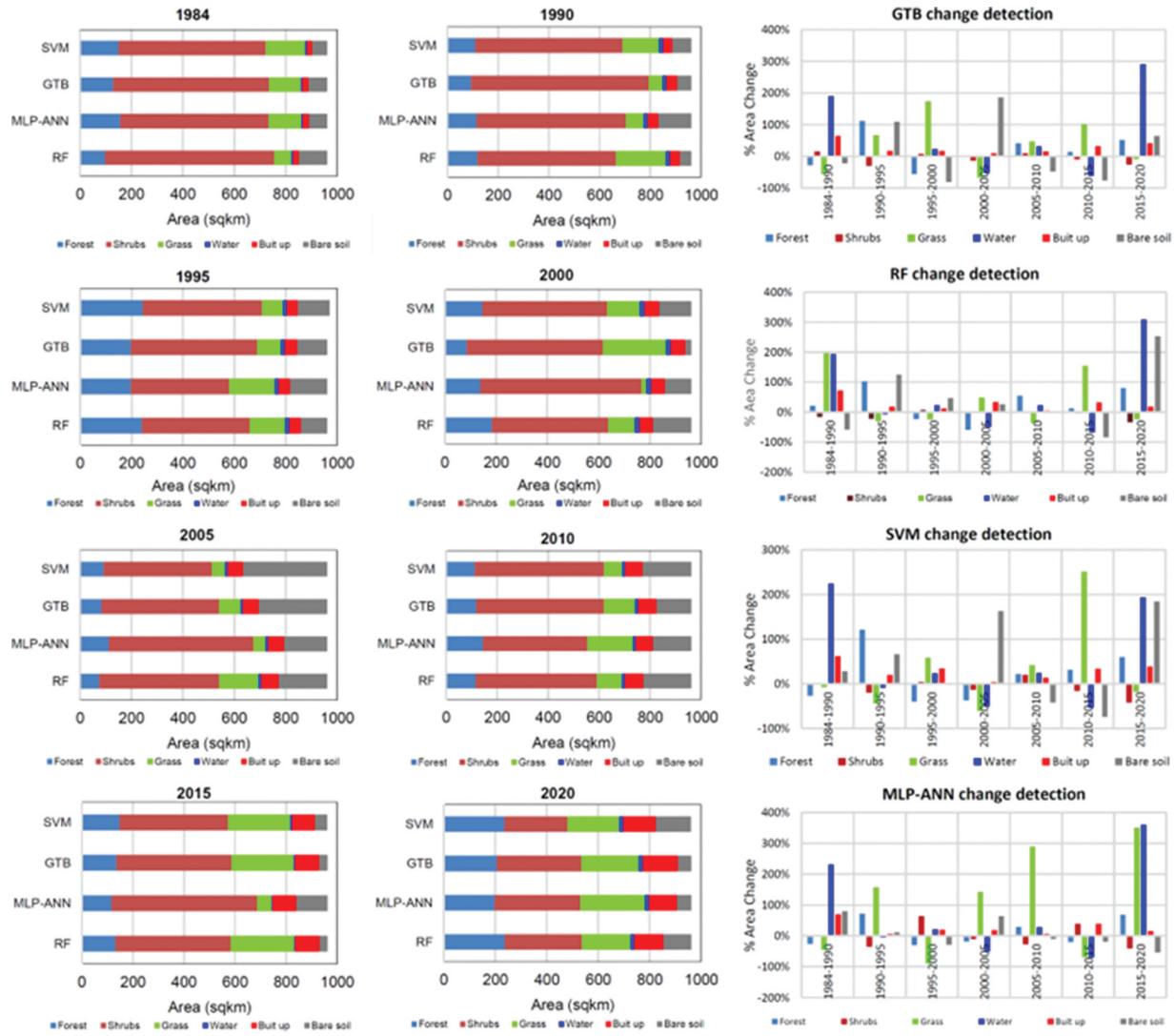


Figure 13. LULC year-class areas and change detection using GTB, RF, SVM and MLP-ANN.

periods of 1984–1990 and 2015–2020 and decreases of more than 50% in 2000–2005 and 2010–2015 (Figure 13).

Bare-soil: Most of the land in the study area are covered by bare soils especially during most parts of the year. For all the years, the classifiers presented different results in terms of the surface area occupied by bare soil/land. In 1984, GTB and MLP-ANN estimated bare soil to occupy 8% (72.35 km^2) of the total area, while RF results indicated that bare soil was 12% (110.82 km^2) and SVM was 6% (57.93 km^2). In 2020, the results from GTB and MLP-ANN were nearly the same at between 5–6% ($52.36\text{--}56.0 \text{ km}^2$), respectively, and RF and SVM estimated bare-soil at 107.70 km^2 (11%) and 137.73 km^2 (14%), respectively. Between 1984 and 2020, the overall area covered by bare soil reduced by 3% (RF), 22% (MLP-ANN), 28% (GTB) and increased by 138% (SVM).

Vegetation: Vegetation cover comprised of grass, shrubs and forested areas. Similar to water, the abundance of natural vegetation cover is influenced by the

climatic conditions. For the 35 years of study, all the classifiers showed that there was an increase in forest and grassland, while the areas covered by shrubs decreased. From RF results, the shrubland reduced by nearly 50% in 1984 from 659.49 km^2 to 300.97 km^2 . In the same duration, MLP-ANN and SVM estimated that shrubs occupied nearly the same area of 578.99 km^2 and 572.66 km^2 , respectively, and in 2020 the classifiers estimated shrubland to be 335.09 km^2 and 246.98 km^2 , respectively. The GTB results indicated that shrubland occupied 607.18 km^2 in 1984 and 331.03 km^2 in 2020 (Figure 13). This implies that most of the land occupied by built areas as a result of conversions from shrublands.

Post-classification feature fusion of classifier results

From the results above, it is observed that an urban LULC class(s) can accurately be detected and extracted by a specific classifier. Thus, to improve on the overall

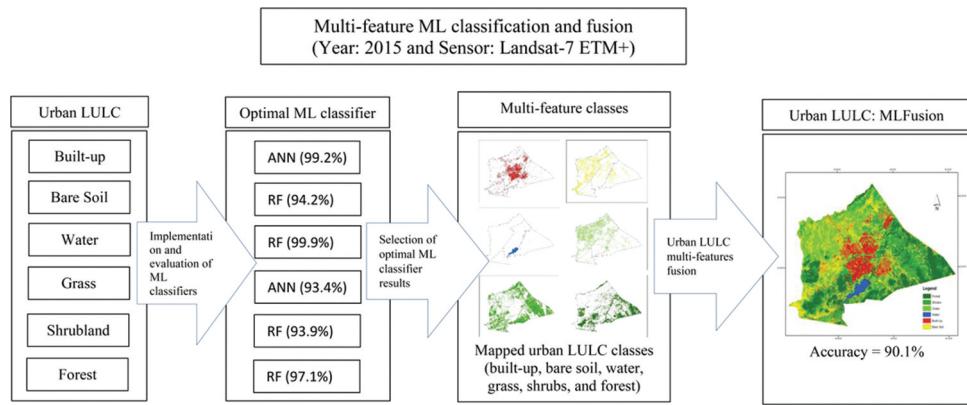


Figure 14. Multiclass FEI-FEO post classification fusion.

accuracy for urban LULC mapping using the ML methods, the proposed FEI-FEO post-classification feature fusion is adopted to maximize on the advantages of the different classifiers and to improve on the accuracy of urban LULC mapping. The output of the ML-fusion results is presented compared in Figure 14, for 2015 which had the least multisensor performance (Figure 12). The results indicate that the proposed ensemble of the best classifier class results as obtained from the MLP-ANN and RF classifiers improves the overall accuracy from 87.5% to 90.1% for 2015. The results of the best classifier class FEI-FEO fusion are compared in Figure 15 with overall improvements in the multitemporal and multisensor urban LULC mapping where the different classes are mapped accurately by different classifiers with results in Figure 12.

Table 5 presents the best class classifier for each year and the resulting post-classification fusion accuracies.

To further illustrate the significance of FEI-FEO approach, Figure 15 presents a comparison of the classification results for 2020 in relation to the ground-truth reference imagery from Google Earth. It is observed that the RF and SVM results have the same shape and structural patterns in terms of mapping the urban areas and captured the bare soils more accurately. In 2020, MLP-ANN tended to underestimate the built-up areas, while GTB classified some urban areas as bare soils. In mapping waterbodies, RF and MLP-ANN were able to differentiate the land–water interfaces better than the other classifiers which tended to map the bare soils around the water bodies as built-up areas. RF detected the shape of the

Table 5. Summary of best class classifier per year.

LULC class	Best machine learning classifier and ML-fusion accuracy (%)							
	1984	1990	1995	2000	2005	2010	2015	2020
Built-up	MLP-ANN	GTB	MLP-ANN	MLP-ANN	MLP-ANN	MLP-ANN	MLP-ANN	RF
Water	MLP-ANN	MLP-ANN	SVM	SVM	MLP-ANN	MLP-ANN	RF	RF
Grass	MLP-ANN	SVM	RF	MLP-ANN	MLP-ANN	RF	MLP-ANN	RF
Shrubs	RF	MLP-ANN	RF	RF	RF	MLP-ANN	RF	RF
Forest	RF	RF	RF	RF	RF	GTB	RF	RF
Bare soil	MLP-ANN	MLP-ANN	RF	SVM	GTB	SVM	RF	RF
ML-Fusion Accuracy	93.7%	93.4%	96.5%	97.1%	95.5%	96.2%	90.1%	93.4%

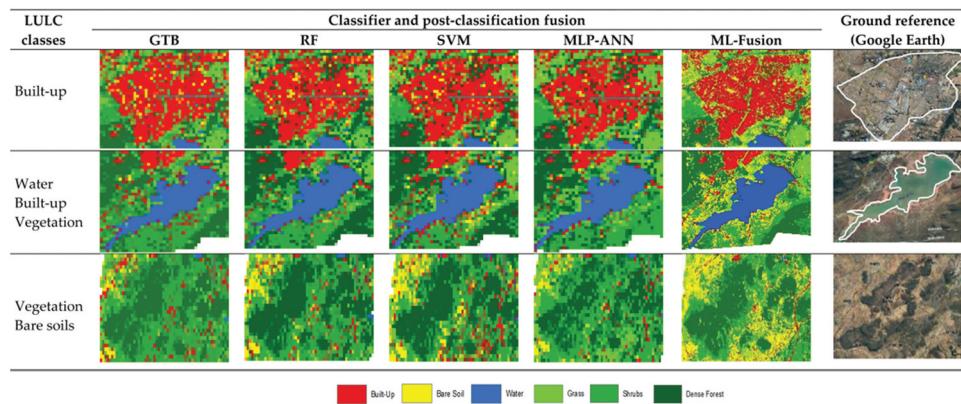


Figure 15. Image-based ground-truth comparison of classification results for different LULC classes and at different locations within the study area for 2020 (reprinted with permission from Y. Ouma et al. (2022). Copyright 2022 ISPRS archives).

dam water body more accurately. For the vegetation cover in 2020, it is observed in [Figure 15](#) that RF and SVM mapped the forest and bare soil areas with the same degree of compactness, while MLP-ANN and GTB tended to map the forest area as mixed with shrubs. Visually, however, the results shows that the classifiers tend to have closely related results.

Discussion

Typified by increasing population and infrastructural development, urbanization is one of the anthropogenic activities that is critical to land-use change. As such, accurate urban LULC information is imperative in providing evidence towards sustainable urban area planning and management. Previous studies have revealed that the capability of classification with remote sensing data, as the most practical data source of urban LULC mapping, is dependent on the classifier, the input data and on the complexity of the landscape ([Klein et al., 2009](#)). On the significance of classifiers in urban LULC mapping, [Pandey et al. \(2021\)](#) noted that the differences in the accuracy resulting from the performance of a classifier was higher than the influence of the land-use land-cover characteristics of a given case study. For mapping complex urban landscapes, the improved performances of machine learning algorithms have resulted in an increase in their applications ([Jozdani et al., 2018](#)). This study thus evaluated the performance of two supervised ensemble classifiers (GTB and RF), pixel-based SVM and neural network MLP machine learning classifiers for urban LULC mapping.

The focus of the pixel-based SVM algorithm is on the pixel independence ([Johnson & Xie, 2013](#)). SVM has been recorded to have advantages that include having high classification accuracy with small training data and is also more robust for data with low noise levels ([Pelletier et al., 2017](#)). On the other hand, the object-based (RF and GTB) and neural network (MLP) classifiers are able to take into account the complex neighborhood spectral and spatial characteristics. Despite several studies having explored the robustness of the performances of different classifiers with remote sensing data sets, the identification of the most appropriate classifier for mapping a specific urban scene and for a specific LULC feature is still a challenging task ([Pandey et al., 2021](#)). Some of the previous studies recorded similar results to this study as represented in [Figure 12](#) and [Table 2](#), where the performances of the classifiers vary for the same scene. However, for a different case study, object-based classifier performed better than pixel-based SVM ([Abdi, 2019](#); [Conchedda et al., 2008](#); [Thanh Noi & Kappas, 2017](#). [Srivastava et al. \(2012\)](#) and [Dixon and Candade \(2008\)](#) compared different machine learning techniques and different Landsat sensors and concluded that

SVM and ANN performed better on Landsat ETM+, while SVM gave better results with Landsat-TM data. These same observations are noted in the current study in which different Landsat sensors MSS, TM, ETM+ and OLI were classified with different accuracy for different years as presented in [Table 4](#). [Huang et al. \(2002\)](#) also noted that with Landsat TM/ETM+, ANN performed better than SVM. [Pal \(2005\)](#) concluded that the performances of RF and SVM were similar if all the required classifier hyperparameters were optimally set. [Erbek et al. \(2004\)](#) and [Deng et al. \(2008\)](#) further reported that the output LULC class areas also varied in the different Landsat sensor satellite data. Besides the scene LULC overall classification accuracy, similar performance patterns are observed in this study for each urban LULC class or feature.

The noted difference between RF and MLP-ANN is that the performance of RF is more stable for all the years and not significantly influenced by the spectral and radiometric differences in the Landsat sensors. The stability of the RF classifier has been reported to be based on its ability to handle category type features, increased number of trees, as well as the bagging and random concepts resulting into its efficiency and precision ([Gislason et al., 2006](#); [Talukdar et al., 2020](#)). Further, the superior performances of RF and MLP-ANN have also been attributed to the fact that the classifiers tend to be tolerant to noise and are significantly more robust towards both random and systematic noise of the training data sets ([Breiman, 2001](#); [Pelletier et al., 2017](#)). Based on its extended feature set ([Georganos et al., 2018](#)), SVM results were observed to be stable with sensor and time and exhibited nearly the same performance trend as RF and MLP-ANN. GTB performance was most effected by the radiometric and spectral resolution differences of the sensors as it recorded the least accuracy. The lower performance by GTB can be attributed to the decision trees being too sensitive to small changes in the training data sets and tends to overfit the model ([Prasad et al., 2006](#)). This implies that the GTB requires continuous readjustments of the of the decision trees to minimize the classification errors ([Awad & Khanna, 2015](#)).

For the study area, MLP-ANN and RF are observed to be the dominant classifiers ([Table 5](#)). Despite the observed marginal differences in the classification results, the study showed that the accuracies of the classifiers were similar at 5% level of significance with all the classifiers performing at above 85% overall accuracy. In a similar evaluation, [Hackman et al. \(2017\)](#) observed that the advanced classification machine learning algorithms may not always be advantageous when applied to process multispectral image data and the focus should be on the abilities of the classifiers to extract specific LULC classes. Despite the standalone classifiers exhibiting good results in

urban LULC classification, more accurate classification techniques are continuously being sought (P. Zhang et al., 2018). Due to the differences in classifier performance for the same scene features, this study proposed the FEI-FEO post-classification feature fusion approach that takes advantage of the classifiers accuracy in mapping a specific urban LULC features. The FEI-FEO hybridization has the potential to discriminate and enhance the robustness and accuracy of urban LULC classification results.

The post-classification results presented statistically in Table 5 and in Figures 12, 14, and comparatively in Figure 15 in terms of classification accuracies shows that the proposed feature fusion approach is suitable for maximizing the performance of machine learning classifiers. The multifeature fusion ensemble takes advantage of the neighborhood and proximities of the pixels mapped in each classifier to generate more accurate and stable urban LULC results. Further, the results in Table 5 with respect to specific class detection indicate the ability of ANN to model nonlinear features and adequately handle the uncertainties that exist in spatial data. The multifeature mapping capability of the proposed FEI-FEO is considered useful in land-use/cover classification tasks in the complex urban environments with high spatial heterogeneity. The utility of FEI-FEO approach should be investigated for each case study to determine the most effective classifiers for mapping specific features.

Conclusions

Mapping of urban landscapes is a complex and challenging task due to the spectral overlaps and spatial heterogeneity of the urban features. This study was carried out, first to implement and evaluate the accuracy of ensemble decision tree-based (GTB and RF) classifiers, and SVM and MLP-ANN machine learning classification algorithms for mapping of urban land-use classes from multitemporal and multisensor Landsat data from 1984 to 2020 at 5-year intervals. The second goal was to maximize the potentials of the ML classifiers for improving the urban LULC mapping through a post-classification feature in – feature out fusion approach. For mapping the six LULC classes over the 35 years, MLP-ANN was the preferred classifier for the urban built-up area and water body. The optimal classifiers for extracting the vegetation classes were determined as grass (MLP-ANN), shrubland (RF) and forest (RF). In terms of the combined overall accuracy for the eight-epoch years, RF average performance was highest at 92.8% which was 1.6%, 1.9% and 5.0% higher than MLP-ANN, SVM and GTB, respectively. For improved ML classifier performances, the study experiments showed that optimal training sample size should be determined. Based on

the advantages of a given classifier in mapping a particular urban LULC class or feature, and to improve on urban LULC mapping accuracy, an ensemble of ML classifiers in recommended in the form of post-classification feature fusion of the best classifier outputs. For multisensor and multitemporal urban LULC mapping, the proposed post-classification FEI-FEO fusion approach increased the accuracy of mapping the urban LULC classes as it combines the inherent feature detection and classification abilities in the machine learning classifiers. The study results show that the detection and mapping of urban LULC classes with a given classifier cannot be generalized and depends on the sensor spectral resolutions, and is also influenced by the temporal, atmospheric, illumination and geometric variabilities. As proposed in this study, accurately derived information on urban LULC is useful for urban land-use planners and managers for decision support on urban planning, management and for growth policy development. This study recommends further evaluations on the performances of the machine learning and the FEI-FO approach in this study in comparison with deep learning classification models for mapping the complex urban environments.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This research project was funded by both the USAID Partnerships for Enhanced Engagement in Research (PEER) under the PEER program cooperative agreement number: AID-OAA-A-11-00012 and the University of Botswana Office of Research and Development (ORD)

ORCID

Yashon O. Ouma  <http://orcid.org/0000-0003-1163-0385>
Jiaguo Qi  <http://orcid.org/0000-0002-8183-0297>

Data availability statement

The data used in this study was obtained from the United States Geological Survey (USGS): <https://earthexplorer.usgs.gov/>. The rest of the data are as presented in this paper. The image data classification was carried out within the Google Earth Engine (GEE).

References

- Abdi, A. M. (2019). Land cover and land use classification performance of machine learning algorithms in a boreal landscape using sentinel-2 data. *GIScience & Remote Sensing*, 57(1), 1–20. <https://doi.org/10.1080/15481603.2019.1650447>

- Awad, M., & Khanna, R. (2015). Support vector machines for classification. In *Efficient learning machines* (pp. 39–66). Apress. https://doi.org/10.1007/978-1-4302-5990-9_3
- Ballanti, L., Blesius, L., Hines, E., & Kruse, B. (2016). Tree species classification using hyperspectral imagery: A comparison of two classifiers. *Remote Sensing*, 8(6), 45. <https://doi.org/10.3390/rs8060445>
- Belgiu, M., & Drăguț, L. (2016). Random forest in remote sensing: A review of applications and future directions. *Isprs Journal of Photogrammetry and Remote Sensing*, 114, 24–31. <https://doi.org/10.1016/j.isprsjprs.2016.01.011>
- Biau, G. Á. Š., & Scornet, E. (2016). A random forest guided tour. *TEST*, 25(2), 197–227. <https://doi.org/10.1007/s11749-016-0481-7>
- Blaschke, T., Hay, G. J., Kelly, M., Lang, S., Hofmann, P., Addink, E., Queiroz Feitosa, R., van der Meer, F., van der Werff, H., van Coillie, F., & Tiede, D. (2014). Geographic object-based image analysis – towards a new paradigm. *Isprs Journal of Photogrammetry and Remote Sensing*, 87, 180–191. <https://doi.org/10.1016/j.isprsjprs.2013.09.014>
- Breiman, L. (2001). Random forests. *Machine Learning*, 45 (1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Camargo, F. F., Sano, E. E., Almeida, C. M., Mura, J. C., & Almeida, T. (2019). A comparative assessment of machine-learning techniques for land use and land cover classification of the Brazilian tropical savanna using ALOS-2/PALSAR-2 polarimetric images. *Remote Sensing*, 11(13), 1600. <https://doi.org/10.3390/rs11131600>
- Carranza-García, M., García-Gutiérrez, J., & Riquelme, J. C. (2019). A framework for evaluating land use and land cover classification using convolutional neural networks. *Remote Sensing*, 11(3), 274. <https://doi.org/10.3390/rs11030274>
- Chen, B., Huang, B., & Xu, B. (2017). Multi-source remotely sensed data fusion for improving land cover classification. *Isprs Journal of Photogrammetry and Remote Sensing*, 124, 27–39. <https://doi.org/10.1016/j.isprsjprs.2016.12.008>
- Conchedda, G., Durieux, L., & Mayaux, P. (2008). An object-based method for mapping and change analysis in mangrove ecosystems. *Isprs Journal of Photogrammetry and Remote Sensing*, 63(5), 578–589. <https://doi.org/10.1016/j.isprsjprs.2008.04.002>
- Dasarathy, B. V. (1997). Sensor fusion potential exploitation-innovative architectures and illustrative applications. Proceedings of the IEEE, 85 (1), 24–38. IEEE. <https://doi.org/10.1109/5.554206>
- Deng, J. S., Wang, K., Deng, Y. H., & Qi, G. J. (2008). PCA-based land-use change detection and analysis using multitemporal and multisensor satellite data. *International Journal of Remote Sensing*, 29(16), 4823–4838. <https://doi.org/10.1080/01431160801950162>
- Dixon, B., & Candade, N. (2008). Multispectral landuse classification using neural networks and support vector machines: One or the other, or both? *International Journal of Remote Sensing*, 29(4), 1185–1206. <https://doi.org/10.1080/01431160701294661>
- Drăguț, L., Tiede, D., & Levick, S. R. (2010). ESP: A tool to estimate scale parameter for multiresolution image segmentation of remotely sensed data. *International Journal of Geographical Information Science*, 24(6), 859–871. <https://doi.org/10.1080/13658810903174803>
- Dwivedi, R. S., Sreenivas, K., & Ramana, K. V. (2005). Land-use/land-cover change analysis in part of Ethiopia using landsat thematic mapper data. *International Journal of Remote Sensing*, 26(7), 1285–1287. <https://doi.org/10.1080/01431160512331337763>
- Erbek, F. S., Özkan, C., & Taberner, M. (2004). Comparison of maximum likelihood classification method with supervised artificial neural network algorithms for land use activities. *International Journal of Remote Sensing*, 25(9), 1733–1748. <https://doi.org/10.1080/0143116031000150077>
- Fan, F., Weng, Q., & Wang, Y. (2007). Land use land cover change in Guangzhou, China, from 1998 to 2003, based on landsat TM/ETM+ imagery. *Sensors*, 7(7), 1323–1342. <https://doi.org/10.3390/s7071323>
- Friedl, M. A., & Brodley, C. E. (1997). Decision tree classification of land cover from remotely sensed data. *Remote sensing of environment*, 61(3), 399–409. [https://doi.org/10.1016/S0034-4257\(97\)00049-7](https://doi.org/10.1016/S0034-4257(97)00049-7)
- Friedman, J. H. (2002). Stochastic gradient boosting. *Computational Statistics & Data Analysis*, 38(4), 367–378. [https://doi.org/10.1016/S0167-9473\(01\)00065-2](https://doi.org/10.1016/S0167-9473(01)00065-2)
- Georganos, S., Grippo, T., Vanhuyse, S., Lennert, M., Shimoni, M., & Wolff, E. (2018). Very high resolution object-based land use-land cover urban classification using extreme gradient boosting. *IEEE Geoscience and Remote Sensing Letters*, 15(4), 607–611. <https://doi.org/10.1109/LGRS.2018.2803259>
- Gislason, P. O., Benediktsson, J. A., & Sveinsson, J. R. (2006). Random forests for land cover classification. *Pattern Recognition Letters*, 27(4), 294–300. <https://doi.org/10.1016/j.patrec.2005.08.011>
- Gong, P., Chen, B., Li, X., Liu, H., Wang, J., Bai, Y., Chen, J., Chen, X., Fang, L., Feng, S., Feng, Y., Gong, Y., Gu, H., Huang, H., Huang, X., Jiao, H., Kang, Y., Lei, G., Li, A., ... Xu, B. (2020). Mapping essential urban land use categories in China (EULUC-China): Preliminary results for 2018. *Science Bulletin*, 65(3), 182–187. <https://doi.org/10.1016/j.scib.2019.12.007>
- Hackman, K. O., Gong, P., & Wang, J. (2017). New land-cover maps of Ghana for 2015 using landsat 8 and three popular classifiers for biodiversity assessment. *International Journal of Remote Sensing*, 38(14), 4008–4021. <https://doi.org/10.1080/01431161.2017.1312619>
- Halder, A., Ghosh, A., & Ghosh, S. (2011). Supervised and unsupervised landuse map generation from remotely sensed images using ant based systems. *Applied Soft Computing*, 11(8), 5770–5781. <https://doi.org/10.1016/j.asoc.2011.02.030>
- Ha, T. V., Tuohy, M., Irwin, M., & Tuan, P. V. Monitoring and mapping rural urbanization and land use changes using landsat data in the northeast subtropical region of Vietnam. (2020). *The Egyptian Journal of Remote Sensing and Space Science*, 23(1), 11–19. U. <https://doi.org/10.1016/j.ejrs.2018.07.001>
- Huang, C., Davis, L., & Townshend, J. (2002). An assessment of support vector machines for land cover classification. *International Journal of Remote Sensing*, 23(4), 725–749. <https://doi.org/10.1080/01431160110040323>
- Jamali, A. (2019). Evaluation and comparison of eight machine learning models in land use/land cover mapping using landsat 8 OLI: A case study of the northern region of Iran. *SN Applied Sciences*, 1(11), 1448. <https://doi.org/10.1007/s42452-019-1527-8>
- Johnson, B., & Xie, Z. (2013). Classifying a high resolution image of an urban area using super-object information. *Isprs Journal of Photogrammetry and Remote Sensing*, 83, 40–49. <https://doi.org/10.1016/j.isprsjprs.2013.05.008>

- Jozdani, S. E., Momeni, M., Johnson, B. A., & Sattari, M. (2018). A regression modelling approach for optimizing segmentation scale parameters to extract buildings of different sizes. *International Journal of Remote Sensing*, 39(3), 684–703. <https://doi.org/10.1080/01431161.2017.1390273>
- Klein, D., Esch, T., Himmller, V., Thiel, M., & Dech, S. (2009). Assessment of urban extent and imperviousness of cape town using TerraSAR-X and landsat images. In Proceedings of the 2009 IEEE International Geoscience and Remote Sensing Symposium, Cape Town, South Africa, 12–17 July. 2009; Volume 3, p. III 1051.
- Knorn, J., Rabe, A., Radeloff, V. C., Kuemmerle, T., Kozak, J., & Hostert, P. (2009). Land cover mapping of large areas using chain classification of neighboring landsat satellite images. *Remote sensing of environment*, 113 (5), 957–964. <https://doi.org/10.1016/j.rse.2009.01.010>
- Kranjčić, N., Medak, D., Župan, R., & Rezo, M. (2019). Machine learning methods for classification of the green infrastructure in city areas. *ISPRS International Journal of Geo-Information*, 8(10), 463. <https://doi.org/10.3390/ijgi8100463>
- Lefulebe, B. E., Van der Walt, A., & Xulu, S. (2022). Fine-scale classification of urban land use and land cover with planetscope imagery and machine learning strategies in the city of Cape Town, South Africa. *Sustainability*, 14 (15), 9139. <https://doi.org/10.3390/su14159139>
- Li, X., Chen, W., Cheng, X., & Wang, L. (2016). A comparison of machine learning algorithms for mapping of complex surface-mined and agricultural landscapes using ZiYuan-3 stereo satellite imagery. *Remote Sensing*, 8(6), 514. <https://doi.org/10.3390/rs8060514>
- Lu, D., & Weng, Q. (2004). Spectral mixture analysis of the urban landscapes in Indianapolis with landsat ETM+ imagery. *Photogrammetric Engineering and Remote Sensing*, 70(9), 1053–1062. <https://doi.org/10.14358/PERS.70.9.1053>
- Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., & Johnson, B. A. (2019). Deep learning in remote sensing applications: A meta-analysis and review. *Isprs Journal of Photogrammetry and Remote Sensing*, 152, 166–177. <https://doi.org/10.1016/j.isprsjprs.2019.04.015>
- Mao, W., Lu, D., Hou, L., Liu, X., & Yue, W. (2020). Comparison of machine-learning methods for urban land-use mapping in Hangzhou City, China. *Remote Sensing*, 12(17), 2817. <https://doi.org/10.3390/rs12172817>
- Maxwell, A. E., Warner, T. A., & Fang, F. (2018). Implementation of machine-learning classification in remote sensing: An applied review. *International Journal of Remote Sensing*, 39(9), 2784–2817. <https://doi.org/10.1080/01431161.2018.1433343>
- Melgani, F., & Bruzzone, L. (2004). Classification of hyperspectral remote sensing images with support vector machines. *IEEE Transactions on Geoscience and Remote Sensing*, 42(8), 1778–1790. <https://doi.org/10.1109/TGRS.2004.831865>
- Mohtadifar, M., Cheffena, M., & Pourafzal, A. (2022). Acoustic- and radio-frequency-based human activity recognition. *Sensors*, 22(9), 3125. <https://doi.org/10.3390/s22093125>
- Myint, S. W., Gober, P., Brazel, A., Grossman-Clarke, S., & Weng, Q. (2011). Per-pixel vs. object-based classification of urban land cover extraction using high spatial resolution imagery. *Remote sensing of environment*, 115(5), 1145–1161. <https://doi.org/10.1016/j.rse.2010.12.017>
- Naidoo, L., Cho, M. A., Mathieu, R., & Asner, G. (2012). Classification of savanna tree species, in the greater kruger national park region, by integrating hyperspectral and LiDAR data in a random forest data mining environment. *Isprs Journal of Photogrammetry and Remote Sensing*, 69, 167–179. <https://doi.org/10.1016/j.isprsjprs.2012.03.005>
- Nery, T., Sadler, R., Solis-Aulestia, M., White, B., Polyakov, M., & Chalak, M. (2016). Comparing supervised algorithms in land use and land cover classification of a landsat time-series. In Proceedings of the International Geoscience and Remote Sensing Symposium (IGARSS), Beijing, China, 3 November. 2016.
- Nevalainen, O., Honkavaara, E., Tuominen, S., Viljanen, N., Hakala, T., Yu, X., Hyppä, J., Saari, H., Pölönen, I., Imai, N., & Tommaselli, A. (2017). Individual tree detection and classification with UAV-based photogrammetric point clouds and hyperspectral imaging. *Remote Sensing*, 9(3), 185. <https://doi.org/10.3390/rs9030185>
- Nichols, J. A., Herbert Chan, H. W., & Baker, M. (2019). Machine learning: Applications of artificial intelligence to imaging and diagnosis. *Biophysical Reviews*, 11(1), 111–118. <https://doi.org/10.1007/s12551-018-0449-9>
- Noble, W. S. (2006). What is a support vector machine? *Nature Biotechnology*, 24(12), 1565–1567. <https://doi.org/10.1038/nbt1206-1565>
- Orieschnig, C. A., Belaud, G., Venot, J. -P., Massuel, S., & Ogilvie, A. (2021). Input imagery, classifiers, and cloud computing: Insights from multi-temporal LULC mapping in the Cambodian Mekong Delta. *European Journal of Remote Sensing*, 54(1), 398–416. <https://doi.org/10.1080/22797254.2021.1948356>
- Ouma, Y. O., Moalafhi, D. B., Anderson, G., Nkwa, B., Odirile, P., Parida, B. P., & Qi, J. (2023). Dam water level prediction using vector autoregression, random forest regression and MLP-ANN models based on land-use and climate factors. *Sustainability*, 14(22), 14934. <https://doi.org/10.3390/su142214934>
- Ouma, Y., Nkwa, B., Moalafhi, D., Odirile, P., Parida, B., Anderson, G., & Qi, J. (2022). Comparison of machine learning classifiers for multitemporal and multisensor mapping of urban LULC features. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLIII-B3-2022, 681–689. <https://doi.org/10.5194/isprs-archives-XLIII-B3-2022-681-2022>
- Pal, M. (2005). Random forest classifier for remote sensing classification. *International Journal of Remote Sensing*, 26 (1), 217–222. <https://doi.org/10.1080/01431160412331269698>
- Pandey, P. C., Koutsias, N., Petropoulos, G. P., Srivastava, P. K., & Ben Dor, E. (2021). Land use/land cover in view of earth observation: Data sources, input dimensions, and classifiers—a review of the state of the art. *Geocarto international*, 36(9), 957–988. <https://doi.org/10.1080/10106049.2019.1629647>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., & Vanderplas, J. (2011). Scikit-learn: Machine learning in python. *Journal of Machine Learning Research: JMLR*, 12(85), 2825–2830. <https://doi.org/10.48550/arXiv.1201.0490>
- Pelletier, C., Valero, S., Ingla, J., Champion, N., Marais Sicr, C., & Dedieu, G. (2017). Effect of training class label noise on classification performances for land cover mapping with satellite image time series. *Remote Sensing*, 9(2), 173. <https://doi.org/10.3390/rs9020173>
- Prasad, A., Iverson, L., & Liaw, A. (2006). Newer classification and regression tree techniques: Bagging and random

- forests for ecological prediction. *Ecosystems (New York, NY)*, 9(2), 181–199. <https://doi.org/10.1007/s10021-005-0054-1>
- Qian, Y., Zhou, W., Yan, J., Li, W., & Han, L. (2015). Comparing machine learning classifiers for object-based land cover classification using very high resolution imagery. *Remote Sensing*, 7(1), 153–168. <https://doi.org/10.3390/rs70100153>
- Rogan, J., Franklin, J., Stow, D., Miller, J., Woodcock, C., & Roberts, D. (2008). Mapping land-cover modifications over large areas: A comparison of machine learning algorithms. *Remote sensing of environment*, 112(5), 2272–2283. <https://doi.org/10.1016/j.rse.2007.10.004>
- Shih, H. C., Stow, D. A., & Tsai, Y. H. (2019). Guidance on and comparison of machine learning classifiers for landsat-based land cover and land use mapping. *International Journal of Remote Sensing*, 40(4), 1248–1274. <https://doi.org/10.1080/01431161.2018.1524179>
- Shi, Y., Qi, Z., Liu, X., Niu, N., & Zhang, H. (2019). Urban land use and land cover classification using multisource remote sensing images and social media data. *Remote Sensing*, 11(22), 2719. <https://doi.org/10.3390/rs11222719>
- Srivastava, P. K., Han, D., Rico-Ramirez, M. A., Bray, M., & Islam, T. (2012). Selection of classification techniques for land use/land cover change investigation. *Advances in Space Research*, 50(9), 1250–1265. <https://doi.org/10.1016/j.asr.2012.06.032>
- Talukdar, S., Singha, P., Praveen, S., Mahato, B., & Rahman, A. (2020). Dynamics of ecosystem services (ESs) in response to land use land cover (LU/LC) changes in the lower gangetic plain of India. *Ecological indicators*, 112, 106121. <https://doi.org/10.1016/j.ecolind.2020.106121>
- Teluguntla, P., Thenkabail, P. S., Oliphant, A., Xiong, J., Gumma, M. K., Congalton, R. G., & Huete, A. (2018). A 30-m landsat-derived cropland extent product of Australia and China using random forest machine learning algorithm on google earth engine cloud computing platform. *Isprs Journal of Photogrammetry and Remote Sensing*, 144, 325–340. <https://doi.org/10.1016/j.isprsjprs.2018.07.017>
- Thanh Noi, P., & Kappas, M. (2017). Comparison of random forest, k-nearest neighbor, and support vector machine classifiers for land cover classification using sentinel-2 imagery. *Sensors*, 18(2), 18. <https://doi.org/10.3390/s18010018>
- Wang, J., Bretz, M., Dewan, M. A. A., & Delavar, M. A. (2022). Machine learning in modelling land-use and land cover-change (LULCC): Current status, challenges and prospects. *The Science of the Total Environment*, 822, 153559. <https://doi.org/10.1016/j.scitotenv.2022.153559>
- Waske, B., & Braun, M. (2009). Classifier ensembles for land cover mapping using multitemporal SAR imagery. *Isprs Journal of Photogrammetry and Remote Sensing*, 64(5), 450–457. <https://doi.org/10.1016/j.isprsjprs.2009.01.003>
- Wu, L., Zhu, X., Lawes, R., Dunkerley, D., & Zhang, H. (2019). Comparison of machine learning algorithms for classification of LiDAR points for characterization of canola canopy structure. *International Journal of Remote Sensing*, 40(15), 5973–5991. <https://doi.org/10.1080/01431161.2019.1584929>
- Yang, C., Wu, G., Ding, K., Shi, T., Li, Q., & Wang, J. (2017). Improving land use/land cover classification by integrating pixel unmixing and decision tree methods. *Remote Sensing*, 9(12), 1222. <https://doi.org/10.3390/rs9121222>
- Yu, L., Liang, L., Wang, J., Zhao, Y., Cheng, Q., Hu, L., Liu, S., Yu, L., Wang, X., Zhu, P., Li, X., Xu, Y., Li, C., Fu, W., Li, X., Li, W., Liu, C., Cong, N., Zhang, H. . . . Gong, P. (2014). Meta-discoveries from a synthesis of satellite-based land-cover mapping research. *International Journal of Remote Sensing*, 35(13), 4573–4588. <https://doi.org/10.1080/01431161.2014.930206>
- Yu, X., Liang, S., & Baaovic, V. (2004). EV-SVM approach for real-time hydrologic forecasting. *Journal of Hydroinformatics*, 6(3), 209–223. <https://doi.org/10.2166/hydro.2004.0016>
- Zhang, P., Ke, Y., Zhang, Z., Wang, M., Li, P., & Zhang, S. (2018). Urban land use and land cover classification using novel deep learning models based on high spatial resolution satellite imagery. *Sensors*, 18(11), 3717. <https://doi.org/10.3390/s18113717>
- Zhang, C., Sargent, I., Pan, X., Li, H., Gardiner, A., Hare, J., & Atkinson, P. M. (2019). Joint deep learning for land cover and land use classification. *Remote sensing of environment*, 221, 173–187. <https://doi.org/10.1016/j.rse.2018.11.014>
- Zong, L., He, S., Lian, J., Bie, Q., Wang, X., Dong, J., & Xie, Y. (2020). Detailed mapping of urban land use based on multi-source data: A case study of Lanzhou. *Remote Sensing*, 12(12), 1987. <https://doi.org/10.3390/rs12121987>